

## SEARCH REQUEST FORM

50597

Scientific and Technical Information Center

Requester's Full Name: \_\_\_\_\_ Examiner #: \_\_\_\_\_ Date: \_\_\_\_\_  
 Art Unit: \_\_\_\_\_ Phone Number 30 \_\_\_\_\_ Serial Number: \_\_\_\_\_  
 Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

*\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

## STAFF USE ONLY

Searcher: \_\_\_\_\_  
 of Contact: Sheppard

Searcher Phone #: \_\_\_\_\_  
 Tel: 308-4499

Searcher Location: \_\_\_\_\_

Date Searcher Picked Up: \_\_\_\_\_

Date Completed: 9/12/07

Searcher Prep & Review Time: \_\_\_\_\_

Clerical Prep Time: \_\_\_\_\_

Online Time: \_\_\_\_\_

## Type of Search

NA Sequence (#) \_\_\_\_\_

AA Sequence (#) \_\_\_\_\_

Structure (#) \_\_\_\_\_

Bibliographic \_\_\_\_\_

Litigation \_\_\_\_\_

Fulltext \_\_\_\_\_

Patent Family \_\_\_\_\_

Other \_\_\_\_\_

## Vendors and cost where applicable

STN \_\_\_\_\_

Dialog \_\_\_\_\_

Questel/Orbit \_\_\_\_\_

Dr. Link \_\_\_\_\_

Lexis/Nexis \_\_\_\_\_

Sequence Systems \_\_\_\_\_

WWW/Internet \_\_\_\_\_

Other (specify) \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**





FT NON\_TER 1 1  
 FT NON\_TER 171 171  
 SQ SEQUENCE 171 AA; 18933 MW; 81BE2EE3FC2C8E72 CRC64;

Query Match 82.1%; Score 284; DB 11; Length 171;  
 Best Local Similarity 79.4%; Pred. No. 4.1e-27;  
 Matches 50; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

OY 1 ADRAAPVIVNLKDELLFPSEALFSGSEGLPKRGARIFSGDKDYLRIHPTWPKSVWHS 60  
 DB 56 ADRSSVPIVNLKDEVLPSPWDLFSGSQGLHSGARIFSGDRDYLRIHPAMPKSVWHS 115

OY 61 DPN 63  
 DB 116 DPS 118

RESULT 5  
 Q90ZD2 PRELIMINARY; PRT; 226 AA.

AC 090ZD2; PRELIMINARY; PRT; 226 AA.  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE COLLAGEN XVIII (FRAGMENT).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY;  
 RA Chen L., Perletti G., Folkman J.;  
 RT "Antitumor activity of rat endostatin."  
 RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF189709; AAF00975.1; -;  
 DR HSP; P39061; IKOE.  
 KW Collagen.  
 FT NON\_TER  
 SQ SEQUENCE 226 AA; 25350 MW; 38B83C0486C0E949 CRC64;

Query Match 82.1%; Score 284; DB 11; Length 226;  
 Best Local Similarity 79.4%; Pred. No. 5.6e-27;  
 Matches 50; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

OY 1 ADRAAPVIVNLKDELLFPSEALFSGSEGLPKRGARIFSGDKDYLRIHPTWPKSVWHS 60  
 DB 106 ADRSSVPIVNLKDEVLPSPWDLFSGSQGLHSGARIFSGDRDYLRIHPAMPKSVWHS 165

OY 61 DPN 63  
 DB 166 DPS 168

RESULT 6  
 O93419 PRELIMINARY; PRT; 386 AA.

AC 093419; PRELIMINARY; PRT; 386 AA.  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE COLLAGEN XVIII (FRAGMENT).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCBI\_TaxId=9031;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RA Halfter W., Dong S., Schurer B., Cole G.;  
 RT "Collagen XVIII is a basement membrane heparan sulfate proteoglycan.";

RL J. Biol. Chem. 0:0-0(1998).  
 DR EMBL; AF083440; AAC33294.1; -;  
 DR HSP; P39061; IKOE.  
 DR InterPro; IPR000087; -;  
 FT NON\_TER  
 SQ SEQUENCE 386 AA; 41775 MW; 34D40FA09EBA3B0E CRC64;

Query Match 73.7%; Score 255; DB 13; Length 386;  
 Best Local Similarity 72.1%; Pred. No. 3.6e-23;  
 Matches 44; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

OY 1 ADRAAPVIVNLKDELLFPSEALFSGSEGLPKRGARIFSGDKDYLRIHPTWPKSVWHS 60  
 DB 266 ADRAVPIVNLKDEVLPSPWDLFSGSEGLPKRGARIFSGDRDYLRIHPAMPKSVWHS 325

OY 61 D 61  
 DB 326 D 326

RESULT 7  
 O35206 PRELIMINARY; PRT; 1367 AA.

AC 035206; PRELIMINARY; PRT; 1367 AA.  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE PROCOLLAGEN, TYPE XV (TYPE XV COLLAGEN).  
 GN COL15A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97480713; PubMed=9339358;  
 RA Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,  
 RA Pihlajaniemi T.;  
 RT "Cloning of mouse type XV collagen sequences and mapping of the  
 RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1  
 RT (XV) collagen sequences indicates divergence in the number of small  
 RT collagenous domains."  
 RL Genomics 45:31-41(1997).  
 RL EMBL; AF011450; AAC53387.1; -;  
 DR HSP; P39061; IKOE.  
 DR MGD; MGI:88449; Col15a1.  
 DR InterPro; IPR000087; -;  
 DR InterPro; IPR001791; -;  
 DR InterPro; IPR003129; -;  
 DR Pfam; PF01391; Collagen; 4.  
 DR Pfam; PF02210; TSPN; 1.  
 DR SMART; SM00282; LamG; 1.  
 SQ SEQUENCE 1367 AA; 140524 MW; A483A1254AF3AECC CRC64;

Query Match 56.9%; Score 197; DB 11; Length 1367;  
 Best Local Similarity 52.4%; Pred. No. 1.9e-15;  
 Matches 33; Conservative 14; Mismatches 16; Indels 0; Gaps 0;

OY 1 ADRAAPVIVNLKDELLFPSEALFSGSEGLPKRGARIFSGDKDYLRIHPTWPKSVWHS 60  
 DB 1247 AERFGLPIVNLKQVLFNNWDSIFSDGGQFNTHIPISFGDRDVTDPSPKSVWHS 1306

OY 61 DPN 63  
 DB 1307 NPH 1309

RESULT 8  
 O9E0D9 PRELIMINARY; PRT; 1367 AA.  
 ID O9E0D9  
 AC O9E0D9;

```

DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, last annotation update)
DE TYPE XV COLLAGEN.
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV.
RX MEDLINE=97480713; PubMed=9339358;
RA Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA Philajaniemi T.;
RT "Cloning of mouse type XV collagen sequences and mapping of the
RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
RT (XV) collagen sequences indicates divergence in the number of small
RT collagenous domains."
RL Genomics 45:31-41(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV.
RX PubMed=11068203;
RA Eklund L., Muona A., Lletard J., Philajaniemi T.;
RT "Structure of the mouse type XV collagen gene, COL15a1, comparison
RT of both genes."
RL Matrix Biol. 19:489-500(2000).
DR EMBL; AF261131; AACG7545.1; JOINED.
DR EMBL; AF261109; AACG7545.1; JOINED.
DR EMBL; AF261110; AACG7545.1; JOINED.
DR EMBL; AF261111; AACG7545.1; JOINED.
DR EMBL; AF261112; AACG7545.1; JOINED.
DR EMBL; AF261113; AACG7545.1; JOINED.
DR EMBL; AF261114; AACG7545.1; JOINED.
DR EMBL; AF261115; AACG7545.1; JOINED.
DR EMBL; AF261116; AACG7545.1; JOINED.
DR EMBL; AF261117; AACG7545.1; JOINED.
DR EMBL; AF261118; AACG7545.1; JOINED.
DR EMBL; AF261119; AACG7545.1; JOINED.
DR EMBL; AF261120; AACG7545.1; JOINED.
DR EMBL; AF261121; AACG7545.1; JOINED.
DR EMBL; AF261122; AACG7545.1; JOINED.
DR EMBL; AF261123; AACG7545.1; JOINED.
DR EMBL; AF261124; AACG7545.1; JOINED.
DR EMBL; AF261125; AACG7545.1; JOINED.
DR EMBL; AF261126; AACG7545.1; JOINED.
DR EMBL; AF261127; AACG7545.1; JOINED.
DR EMBL; AF261128; AACG7545.1; JOINED.
DR EMBL; AF261129; AACG7545.1; JOINED.
DR EMBL; AF261130; AACG7545.1; JOINED.
RW COLLAGEN.
SO SEQUENCE 1367 AA; 140454 MW; 40F259A3CD47C982 CRC64;

Query Match 56.68; Score 196; DB 11; Length 1367;
Best Local Similarity 53.28; Pred. No. 2,5e-15;
Matches 33; Conservative 13; Mismatches 16; Indels 0; Gaps 0;
QY 1 ADRAAVPIVNLKDELLFSPSWALFSGSEGLPKPGARIFSPFGKDYLRPTWPOKSVWHGS 60
DB 1247 AERGLPLIVNLKGVLFNNMDSIFSGDGGQFNTHPIYISFGGRVMDPSPPOKVVWHGS 1306
QY 61 DP 62
DB 1307 NP 1308

RESULT 9
QY14M4 PRELIMINARY; PRT; 1388 AA.
AC Q9Y4M4;

```

```

DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, last annotation update)
DE TYPE XV COLLAGEN.
GN COL15A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=94148920; PubMed=8106446;
RA Kivirikko S., Heinamaki P., Rehn M., Honkanen N., Myers J.C.,
RA Philajaniemi T.;
RT "Primary structure of the alpha 1 chain of human type XV collagen and
RT exon-intron organization in the 3' region of the corresponding gene."
RL J. Biol. Chem. 269:4773-4779(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=98316357; PubMed=9651385;
RA Hagg P.M., Muona A., Lletard J., Kivirikko S., Philajaniemi T.;
RT "Complete exon-intron organization of the human gene for the alpha 1
RT chain of type XV collagen (COL15A1) and comparison with the homologous
RT COL18A1 gene."
RL J. Biol. Chem. 273:17824-17831(1998).
DR EMBL; L25280; AAC78500.1; JOINED.
DR EMBL; AF052956; AAC78500.1; JOINED.
DR EMBL; AF052957; AAC78500.1; JOINED.
DR EMBL; AF052958; AAC78500.1; JOINED.
DR EMBL; AF052959; AAC78500.1; JOINED.
DR EMBL; AF052960; AAC78500.1; JOINED.
DR EMBL; AF052961; AAC78500.1; JOINED.
DR EMBL; AF052962; AAC78500.1; JOINED.
DR EMBL; AF052963; AAC78500.1; JOINED.
DR EMBL; AF052964; AAC78500.1; JOINED.
DR EMBL; AF052965; AAC78500.1; JOINED.
DR EMBL; AF052966; AAC78500.1; JOINED.
DR EMBL; AF052967; AAC78500.1; JOINED.
DR EMBL; AF052968; AAC78500.1; JOINED.
DR EMBL; AF052969; AAC78500.1; JOINED.
DR EMBL; AF052970; AAC78500.1; JOINED.
DR EMBL; AF052971; AAC78500.1; JOINED.
DR EMBL; AF052972; AAC78500.1; JOINED.
DR EMBL; AF052973; AAC78500.1; JOINED.
DR EMBL; AF052974; AAC78500.1; JOINED.
DR EMBL; AF052975; AAC78500.1; JOINED.
DR EMBL; L25285; AAC78500.1; JOINED.
DR EMBL; L25284; AAC78500.1; JOINED.
DR EMBL; L25283; AAC78500.1; JOINED.
DR EMBL; L25282; AAC78500.1; JOINED.
DR EMBL; L25281; AAC78500.1; JOINED.
DR HSSP; P39061; IKOE.
DR InterPro; IPR000087; -.
DR InterPro; IPR001129; -.
DR Pfam; PF01391; Collagen; 4.
DR SMART; PF02210; TSPN; 1.
RW SMART; SM00282; Lamc; 1.
RW COLLAGEN.
SO SEQUENCE 1388 AA; 141756 MW; 96828E45E847194B CRC64;

Query Match 55.88; Score 193; DB 4; Length 1388;
Best Local Similarity 49.28; Pred. No. 5,9e-15;
Matches 31; Conservative 16; Mismatches 16; Indels 0; Gaps 0;
QY 1 ADRAAVPIVNLKDELLFSPSWALFSGSEGLPKPGARIFSPFGKDYLRPTWPOKSVWHGS 60
DB 1268 AERYSLPIVNLKGVLFNNMDSIFSGHGGQFNTHPIYISFGGRVMDPSPPOKVVWHGS 1327
QY 61 DPN 63
DB 1328 SPH 1330

```

```

RESULT 10
Q9V509 PRELIMINARY: PRT: 581 AA.
ID 09V509:
AC 09V509:
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG8645 PROTEIN.
GN CG8645.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA Sulton G.C., Morten J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Sutton G.C., Worland J.R., White S.E., Richards S., Ashburner M., Henderson S.N.,
RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abill J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borrova D., Botchan M.R., Bouck J., Broxstein P., Brotter P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Daveport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson C., Dong L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.F., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harits N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegawa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K.A., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Splier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Massaman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT *The genome sequence of Drosophila melanogaster.*;
RL Science 287:2185-2195(2000).
DR EMBL: AE003560; AAF50621.1; -
DR HSSP: P39061; IKOE.
DR FLYbase: FBgn0035732; CG8645.
DR InterPro: IPR000087; -
DR InterPro: IPR002088; -
DR Pfam: PF01391; Collagen; 3.
DR PROSITE: PS00904; PPTA; UNKNOWN.1.
SO SEQUENCE 581 AA; 60772 MW; 19EC1E48CBA77EE7 CRC64;

```

```

Query Match 49.0%; Score 169.5; DB 5; Length 581;
Best Local Similarity 49.2%; Pred. No. 1.7e-12;
Matches 31; Conservative 11; Mismatches 20; Indels 1; Gaps 1;
QY 1 ADRAAVPIVNLKDELLFPSSWALFSGSEGPLKPGARLIFSGDKDYLNRHPTWPKSVWHS 60

```

```

DB 429 ADR-DLPVNTRGDVLFPNSKMGIFNGGFFSOAPRTISFGKVMMDSTWPKKMWHS 487
QY 61 DPN 63
DB 488 LPN 490
RESULT 11
ID 017866 PRELIMINARY: PRT: 650 AA.
AC 017866:
DT 01-JUN-1998 (TrEMBLrel. 05, Created)
DT 01-JUN-1998 (TrEMBLrel. 07, Last sequence update)
DE 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE F39H11.4 OR CLE-1.
GN F39H11.4 OR CLE-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodertinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lighning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierrey-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.*;
RL Nature 368:32-38(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA Kramer J.M., Ackley B.D.;
RT Loss of endostatin domain from C. elegans type XVIII collagen
RT homologues causes cell and axon guidance defects.*;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z81079; CAB03084.1; -
DR HSSP: ARI64859; ADA47825.1; -
DR InterPro: IPR000087; -
DR Pfam: PF01391; Collagen; 2.
SO SEQUENCE 650 AA; 69596 MW; 6CF29ED9C16B170E CRC64;

```

```

Query Match 40.5%; Score 140; DB 5; Length 650;
Best Local Similarity 47.2%; Pred. No. 7.7e-09;
Matches 25; Conservative 9; Mismatches 17; Indels 2; Gaps 1;
QY 8 IVNLKDELLFPSSWALFSGSEGPLKPGARLIFSGDKDYLNRHPTWPKSVWHS 60
DB 525 VVNAVGHLPFSWRSFVNGAO--MNPALKLFSFRHDVLDNSRWPDKRWHS 575
RESULT 12
ID 0909K6 PRELIMINARY: PRT: 778 AA.
AC 0909K6:
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CLE-1B PROTEIN.
GN CLE-1.

```

Search completed: September 10, 2001, 13:50:18  
Job time: 108 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2001, 13:48:45 ; Search time 10.07 Seconds

(without alignments)  
214.309 Million cell updates/sec

Title: US-09-822-540A-1

Sequence: 346 1 ADRAVPYVKNKDELLFPSPW.....DYLHPWPQKSVWHGSDPN 63

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	346	100.0	1516	1	CAH1_HUMAN
2	294	85.0	1527	1	CAH1_MOUSE
3	193	55.8	1388	1	CAH1_HUMAN
4	63.5	18.4	266	1	CR21_STNL
5	61	17.6	683	1	ACSL_KULU
6	61	17.6	1043	1	PRTS_SERMA
7	58.5	16.9	1720	1	FTSH_CHLVU
8	57.5	16.6	553	1	FXCI_HUMAN
9	56.5	16.5	770	1	TRPG_ASPNG
10	56.5	16.3	884	1	YP67_MYCTU
11	56.5	16.3	884	1	YP67_MYCTU
12	56	16.2	651	1	SYM_METUA
13	55.5	16.0	727	1	NUAN_HUMAN
14	55.5	16.0	1174	1	KCRF_STRPU
15	55	15.9	339	1	XAP5_HUMAN
16	55	15.9	380	1	OP54_DROPS
17	55	15.9	440	1	YA33_SCHPO
18	55	15.9	444	1	PAPA_CANFA
19	55	15.9	545	1	AIRE_HUMAN
20	55	15.9	683	1	KPCL_RAT
21	54.5	15.8	1861	1	MAP2_RAT
22	54	15.6	215	1	COAT_PMV
23	54	15.6	434	1	KP58_MOUSE
24	54	15.6	436	1	KP58_MOUSE
25	54	15.6	494	1	VATB_PLAFA
26	54	15.6	682	1	KPCL_HUMAN
27	53.5	15.5	705	1	YOG4_CAEL
28	53.5	15.5	436	1	Y450_RHISN
29	53	15.3	436	1	KP58_HUMAN
30	53	15.3	549	1	SVR_ARCFU
31	52.5	15.2	695	1	TKT_RHIME
32	52.5	15.2	74	1	SOM1_YEAST
33	52.5	15.2	317	1	GCH_RAT

34	52.5	15.2	350	1	FLA1_TREPA	P18193	treponema p
35	52.5	15.2	363	1	DCAM_SPIOI	P46255	spiniacola o
36	52.5	15.2	366	1	YBA8_BACCI	P48843	bacillus ci
37	52.5	15.2	372	1	OPRD_MOUSE	P32300	mus musc
38	52.5	15.2	376	1	WTS1_RHIME	O30569	herpes simp
39	52.5	15.2	489	1	VHS_HSV11	P10225	herpes simp
40	52.5	15.2	713	1	ACSL_YEAST	O01574	saccharomy
41	52.5	15.2	808	1	PLD_TOBAC	P93400	nicotiana t
42	52	15.0	267	1	CR22_ARATH	P04778	arabidopsis
43	52	15.0	347	1	FOS_CYPCA	P79702	cyprinus ca
44	52	15.0	407	1	Y116_MYCTU	P72052	mycobacteri
45	52	15.0	415	1	VE2_PABVE	P11329	european el

## ALIGNMENTS

RESULT 1  
ID CAH1\_HUMAN STANDARD: PRT: 1516 AA.  
AC P39060; 09Y608; 09Y607; 09UK38;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE COLLAGEN ALPHA 1(XVIII) CHAIN PRECURSOR [CONTAINS: ENDOSTATIN].  
GN COL18A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid:9606;  
RN [1]  
RP MEDLINE-98164096; PubMed-9503365;  
RX Saarela J., Ylikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.,  
\*Complete primary structure of two variant forms of human type XVIII  
RT collagen and tissue-specific differences in the expression of the  
RL corresponding transcripts.\*;  
RL Matrix Biol. 16:319-328(1998).  
RN [2]  
RP MEDLINE-20289799; PubMed-10830953;  
RX Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,  
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,  
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,  
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,  
RA Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,  
RA Rosenthal A., Knuch J., Shibuya K., Kawasaki K., Asakawa S.,  
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,  
RA Minoshima S., Shimizu N., Nordliek G., Hornischer K., Brandt P.,  
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,  
RA Rammer J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,  
RA Wehnach H., Reinhardt R., Yaspo M.-L., Nizetic D., Francis F.,  
\*The DNA sequence of human chromosome 21.\*;  
RL Nature 405:311-319(2000).  
RN [3]  
RP MEDLINE-94245237; PubMed-8188291;  
RX Oh S.-P., Warman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,  
RA Olsen B.R.,  
\*Cloning of cDNA and genomic DNA encoding human type XVIII collagen  
and localization of the alpha 1(XVIII) collagen gene to mouse  
RT chromosome 10 and human chromosome 21.\*;  
RL Genomics 19:494-499(1994).  
RN [4]  
RP MEDLINE-1334-1516 FROM N.A.  
RX Tissue-Placenta;  
RA Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.,  
\*Cloning and expression of human endostatin gene in Escherichia  
RT coli.\*;  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP INVOLVEMENT IN KNOBLOCH SYNDROME.



```

RN [3]
RP SEQUENCE OF 213-1140 FROM N.A. (SHORT ISOFORM).
RX MEDLINE-94240112; PubMed-8183894;
RA Rehn M.V., Pihlajaniemi T.,
RT "Alpha 1(XVIII), a collagen chain with frequent interruptions in the
RT collagenous sequence, a distinct tissue distribution, and homology
RT with type XV collagen."
RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RN [4]
RP SEQUENCE OF 240-1527 FROM N.A.
RX TISSUE-LIVER;
RX MEDLINE-94240111; PubMed-8183893;
RA Oh S.P., Kamagata Y., Muragaki Y., Timmons S., Ooshima A., Olsen B.R.;
RT "Isolation and sequencing of cDNAs for proteins with multiple domains
RT of gly-xaa-yaa repeats identify a distinct family of collagenous
RT proteins."
RL Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).
RN [5]
RP CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.
RX MEDLINE-97160848; PubMed-9008168;
RA O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vasios G., Lane W.S.,
RA Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;
RT "Endostatin: an endogenous inhibitor of angiogenesis and tumor
RT growth."
RL Cell 88:277-285(1997).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.
RX MEDLINE-98169382; PubMed-9501087;
RA Hohenester E., Sasaki T., Olsen B.R., Timpl R.;
RT "Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A
RT resolution."
RL EMBO J. 17:1656-1664(1998).
CC -1- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL
CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
CC BINDING TO THE HEPARAN SULFATE PROTEGLYCANS INVOLVED IN GROWTH
CC FACTOR SIGNALING.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, A SHORT FORM AND A LONG FORM
CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPLET REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: BELONGS TO THE MULTIPLEXIN FAMILY OF COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L16898; AAA37434.1; -
DR EMBL: U03714; AAA20657.1; -
DR EMBL: U03715; AAC52901.1; -
DR EMBL: U34606; AAC52901.1; JOINED.
DR EMBL: U34608; AAC52901.1; JOINED.
DR EMBL: U34609; AAC52901.1; JOINED.
DR EMBL: U34610; AAC52901.1; JOINED.
DR EMBL: U34611; AAC52901.1; JOINED.
DR EMBL: U34612; AAC52901.1; JOINED.
DR EMBL: U34613; AAC52901.1; JOINED.
DR EMBL: U03716; AAC52901.1; JOINED.
DR EMBL: U03718; AAC52901.1; JOINED.
DR EMBL: U03715; AAC52902.1; -
DR EMBL: U34607; AAC52902.1; JOINED.
DR EMBL: U34609; AAC52902.1; JOINED.
DR EMBL: U34610; AAC52902.1; JOINED.
DR EMBL: U34611; AAC52902.1; JOINED.
DR EMBL: U34612; AAC52902.1; JOINED.
DR EMBL: U34613; AAC52902.1; JOINED.
DR EMBL: U03716; AAC52902.1; JOINED.
DR EMBL: U03718; AAC52902.1; JOINED.
DR EMBL: U03716; AAC52902.1; JOINED.
DR EMBL: U03718; AAC52902.1; JOINED.
DR EMBL: U11636; AAC52178.1; -

```

```

DR EMBL: L22545; AAA19787.1; -
DR PDB: 1KOE; 16-FEB-99.
DR MGD: MGI:88451; Col18a1.
DR InterPro: IPR000087; -
DR Pfam: PF01391; Collagen; 6.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
KW 3D-structure.
FT SIGNAL 1 26
FT CHAIN 27 1527
FT CHAIN 1344 1527
FT DOMAIN 27 1527
FT DOMAIN 539 538
FT DOMAIN 539 535
FT DOMAIN 566 575
FT DOMAIN 576 649
FT DOMAIN 650 673
FT DOMAIN 674 795
FT DOMAIN 796 818
FT DOMAIN 819 901
FT DOMAIN 902 915
FT DOMAIN 916 957
FT DOMAIN 958 970
FT DOMAIN 971 1043
FT DOMAIN 1044 1053
FT DOMAIN 1054 1086
FT DOMAIN 1087 1098
FT DOMAIN 1099 1122
FT DOMAIN 1123 1129
FT DOMAIN 1130 1181
FT DOMAIN 1182 1194
FT DOMAIN 1195 1212
FT DOMAIN 1213 1527
FT CARBOHYD 338 700
FT CARBOHYD 1376 1516
FT DISULFID 1478 1508
FT SITE 1104 1106
FT SITE 1104 1212
FT VARSPLIC 1 238
FT CONFLICT 900 900
FT CONFLICT 947 947
FT CONFLICT 964 964
FT CONFLICT 1157 1157
FT CONFLICT 1266 1266
FT CONFLICT 1276 1276
FT CONFLICT 1437 1437
FT CONFLICT 1437 1437
SQ SEQUENCE 1527 AA; 156008 MM; 9645045AF140B513 CRC64;

CELL ATTACHMENT SITE (POTENTIAL).
MISSING (IN SHORT ISOFORM).
APPTOLPPPOSIALQADLRSPAPD -> MAPRRHLLDVL
TSVLDAVARVSAE (IN SHORT ISOFORM).
P -> L (IN REF. 4).
P -> F (IN REF. 4).
A -> R (IN REF. 4).
R -> P (IN REF. 4).
P -> L (IN REF. 4).
L -> F (IN REF. 4).
L -> V (IN REF. 4).

Query Match 85.08; Score 294; DB 1; Length 1527;
Best Local Similarity 81.0%; Pred. No. 4; 9e-27;
Matches 51; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLPPSEALFSGSEGPLKPCARIFSPDGKVDLRHPTMPKQSVHGS 60
DB 1407 ADRAAVPIVNLKDEVLSPMSDLFSGSQGLQPARIFSPDGRVLRHPTMPKQSVHGS 1466
QY 61 DPN 63
DB 1467 DPS 1469

RESULT 3
CAIE_HUMAN STANDARD; PRT; 1388 AA.
ID CAIE_HUMAN
AC P39059;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 1(XV) CHAIN PRECURSOR.
GN COL15A1.
OS Homo sapiens (Human).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=umbilical cord;  
 RX MEDLINE=94148920; PubMed=8106446;  
 RA Kivirikko S., Helenaki P., Rehn M.V., Honkanen N., Myers J.C.,  
 RA Pihlajaniemi T.;  
 RT "Primary structure of the alpha 1 chain of human type XV collagen and  
 RT exon-intron organization in the 3' region of the corresponding  
 RT gene.";  
 RL J. Biol. Chem. 269:4773-4779(1994).  
 RN [2]  
 RP SEQUENCE OF 1-569 FROM N.A.  
 RC TISSUE=placenta;  
 RX MEDLINE=94140817; PubMed=8307960;  
 RA Murgaki Y., Abe N., Ntombiya Y., Olsen B.R., Ooshima A.;  
 RT "The human alpha 1(XV) collagen chain contains a large amino-terminal  
 RT non-triple helical domain with a tandem repeat structure and homology  
 RT to alpha 1(XVIII) collagen.";  
 RL J. Biol. Chem. 269:4042-4046(1994).  
 RN [3]  
 RP SEQUENCE OF 544-1252 FROM N.A.  
 RX MEDLINE=93061196; PubMed=1279671;  
 RA Myers J.C., Kivirikko S., Gordon M.K., Dion A.S., Pihlajaniemi T.;  
 RT "Identification of a previously unknown human collagen chain, alpha  
 RT 1(XV), characterized by extensive interruptions in the triple-helical  
 RT region.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:10144-10148(1992).  
 RN [4]  
 RP TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN INTERNAL ORGANS  
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN INTERNAL ORGANS  
 CC SUCH AS ADRENAL GLAND, PANCREAS AND KIDNEY.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: BELONGS TO THE MULTIPLEXIN FAMILY OF COLLAGENS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; L25286; AAA58429.1;  
 DR EMBL; D21230; BAA04762.1;  
 DR EMBL; L01697; -; NOT\_ANNOTATED\_CDS.  
 DR MIM; 120325; -;  
 DR InterPro; IPR000087; -;  
 DR Pfam; PF01391; Collagen; 4;  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Cell adhesion; Collagen; Glycoprotein; Signal.  
 FT SIGNAL 1 25  
 FT CHAIN 26 1388  
 FT DOMAIN 26 555  
 FT DOMAIN 556 573  
 FT DOMAIN 574 618  
 FT DOMAIN 619 732  
 FT DOMAIN 733 763  
 FT DOMAIN 764 798  
 FT DOMAIN 799 822  
 FT DOMAIN 823 867  
 FT DOMAIN 868 878  
 FT DOMAIN 879 949  
 FT DOMAIN 950 983  
 FT DOMAIN 984 1013  
 FT DOMAIN 1014 1027  
 FT DOMAIN 1028 1045  
 FT DOMAIN 1046 1052  
 FT DOMAIN 1053 1107  
 FT DOMAIN 1108 1117  
 FT DOMAIN 1118 1132  
 FT DOMAIN 1133 1388  
 FT COLLAGEN ALPHA 1(XV) CHAIN.  
 FT NONHELICAL REGION 1 (NC1).  
 FT TRIPLE-HELICAL REGION 1 (COL1).  
 FT NONHELICAL REGION 2 (NC2).  
 FT TRIPLE-HELICAL REGION 2 (COL2).  
 FT NONHELICAL REGION 3 (NC3).  
 FT TRIPLE-HELICAL REGION 3 (COL3).  
 FT NONHELICAL REGION 4 (NC4).  
 FT TRIPLE-HELICAL REGION 4 (COL4).  
 FT NONHELICAL REGION 5 (NC5).  
 FT TRIPLE-HELICAL REGION 5 (COL5).  
 FT NONHELICAL REGION 6 (NC6).  
 FT TRIPLE-HELICAL REGION 6 (COL6).  
 FT NONHELICAL REGION 7 (NC7).  
 FT TRIPLE-HELICAL REGION 7 (COL7).  
 FT NONHELICAL REGION 8 (NC8).  
 FT TRIPLE-HELICAL REGION 8 (COL8).  
 FT NONHELICAL REGION 9 (NC9).  
 FT TRIPLE-HELICAL REGION 9 (COL9).  
 FT NONHELICAL REGION 10 (NC10).

FT DOMAIN 358 555  
 FT REPEAT 358 408  
 FT REPEAT 409 459  
 FT REPEAT 460 509  
 FT REPEAT 510 555  
 FT CARBOHYD 306 306  
 FT CARBOHYD 324 324  
 FT CARBOHYD 687 687  
 FT CARBOHYD 807 807  
 FT CARBOHYD 814 814  
 FT CARBOHYD 1046 1046  
 FT CONFLICT 10 10  
 FT CONFLICT 49 49  
 FT CONFLICT 95 95  
 FT CONFLICT 150 150  
 FT CONFLICT 204 204  
 FT CONFLICT 409 409  
 SQ SEQUENCE 1388 AA; 141930 MW; 60822AD925A3093D CRC64;  
 Query Match 55.8%; Score 193; DB 1; Length 1388;  
 Best Local Similarity 49.2%; Pred. No. 4.5e-15;  
 Matches 31; Conservative 16; Mismatches 16; Indels 0; Gaps 0;  
 OY 1 ADRAAPVIVNKLDELPSWEALFSGSGPLKPGARIFSGCKDYLRLPTWPKSWHGS 60  
 DB 1268 AERYSLPVLNKGQVLFNNWDSIFSGHGQFMNHIPITFSFGDRIDMTDPSWPKVYHGS 1327  
 OY 61 DPN 63  
 DB 1328 SPH 1330  
 RESULT 4  
 ID CB21\_SINAL STANDARD; PRT; 266 AA.  
 AC P13851;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE CHLOROPHYLL A-B BINDING PROTEIN 1 PRECURSOR (LHCII TYPE I CAB-1)  
 DE (LHCp).  
 GN CAB1.  
 OS Sinapis alba (White mustard) (Brassica hirta).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Sinapis.  
 OX NCBI\_TaxID=3728;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Chloplast;  
 RX MEDLINE=92322955; PubMed=1623179;  
 RA Gailly A., Bartschauer A., von Arnim A., Koesse H.;  
 RT "Isolation and characterization of a gene encoding a chlorophyll a/b-  
 RT binding protein from mustard and the targeting of the encoded protein  
 RT to the thylakoid membrane of pea chloroplasts in vitro.";  
 RL Plant Mol. Biol. 19:277-287(1992).  
 CC -1- FUNCTION: THE LIGHT-HARVESTING COMPLEX (LHC) FUNCTIONS AS A LIGHT  
 CC RECEPTOR. IT CAPTURES & DELIVERS EXCITATION ENERGY TO PHOTOSYSTEMS  
 CC WITH WHICH IT IS CLOSELY ASSOCIATED. THE N-TERMINUS OF THE PROTEIN  
 CC EXTENDS INTO THE STROMA WHERE IT IS INVOLVED WITH ADHESION OF  
 CC GRANA MEMBRANES AND PHOTOREGULATED BY REVERSIBLE PHOSPHORYLATION  
 CC OF ITS THREONINE RESIDUES. BOTH ARE BELIEVED TO MEDIATE THE  
 CC DISTRIBUTION OF EXCITATION ENERGY BETWEEN PHOTOSYSTEMS I AND II.  
 CC -1- SUBUNIT: THE LHC COMPLEX CONSISTS OF CHLOROPHYLLS A & B AND  
 CC CHLOROPHYLL A-B BINDING PROTEINS.  
 CC CHLOROPHYLL A-B BINDING LOCATIONS: CHLOROPLAST THYLAKOID MEMBRANE.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

Query Match	Similarity	18.4%	Score 63.5	DB 1	length 266
Best Local	Similarity	39.0%	Pred. No. 1.9		
Matches	16	Conservative	4	Mismatches	18
				Indels	3
				Gaps	1
24	FGSGEGPLKPGARIFSPDGKDYLR---HPTWQKSYMGSD	61			
1	1	1	1	1	1
13	FAGKAAYKLPSPGAEVFGTGRVTMRKTVKPTGSGSPWYSSD	53			
DB					

	Matches	16;	Conservative	8;	Mismatches	13;	Indels	22;	Gaps	3
QY	25	SGSEPTLRPGARIFESFDKDV-LRHPT-----WPG-KSVNHGSD	61	:	: ::   :	:	:	:	:	:
Db	460	AGCATPMMRPGAAPFFFGDGLAVLDPTTGISIoTGEHAECVALIKRRWPSPFARTIMKNND	518							

Specific excretion of serritin marcescens protease through the outer membrane of *Escherichia coli*."; *J. Bacteriol.* 166:937-944(1986).

[2]  
PARTIAL SEQUENCE, AND PROCESSING. MEDLINE=92348352; PubMed=1639760; Shikata S., Shimada K., Kataoka H., Hornouchi S., Beppu T.; "Detection of large COOH-terminal domains processed from the precursor of *Serratia marcescens* serine protease in the outer membrane of *Escherichia coli*."; *J. Biochem.* 111:627-632(1992).

-I- SUBCELLULAR LOCATION: SECRETED.  
-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE SUBTILASE FAMILY.

	Query Match	17.68;	Score 61;	DB 1;	Length 1045;
	Best Local Similarity	34.98;	Pred. No. 17;		
	Matches 15; Conservative	4;	Mismatches 24;	Indels 0;	Gaps 0
QY	14 ELLEFPSEALFSGSPGLKPGARIFESPDGDLVLRHPWPQSV	56			
	:     :				

DB 449 ECFSDSWNDISGHCGLTKGTAGTLALGNNTYRGDTWQGV 491

RESULT 7  
FISH\_CHLVU STANDARD; PRT; 1720 AA.

ID FISH\_CHLVU  
AC P56369;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE CELL DIVISION PROTEIN FISH HOMOLOG.  
GN FISH.  
OS Chlorella vulgaris.  
OC Chlorophyta: Viridiplantae; Chlorophyta: Trebouxiophyceae; Chlorellales;  
OC Chlorellaceae; Chlorella.  
OX NCBI\_Taxid=3077;

RN (1)  
RC SEQUENCE FROM N.A.  
RC STRAIN-TAM C-27 / TAMIVA.  
RA MEDLINE-97303241; PubMed-9159184;  
RA Matsuura T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,  
RA Tsuzuki J., Nakashima K., Tsuzuki T., Suzuki Y., Hamada A., Ohta T.,  
RA Imanura A., Yoshinaga K., Sugita M.,  
RT "Complete nucleotide sequence of the chloroplast genome from the  
RT green alga Chlorella vulgaris: the existence of genes possibly  
RT involved in chloroplast division."  
RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).  
CC -1 SUBCELLULAR LOCATION: CHLOROPLAST; INTEGRAL MEMBRANE PROTEIN  
CC (POTENTIAL).  
CC -1 DOMAIN: LACKS THE ZINC PROTEASE DOMAIN OF OTHER FISH PROTEINS.  
CC ALSO MUCH LONGER IN BOTH THE N- AND C-TERMINUS.  
CC -1 SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.  
CC  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC  
CC EMBL: AB001684; BAA57905.1; -  
CC InterPro: IPR001939; -  
CC DR Pfam: PF00004; AAA; 1.  
CC DR PROSITE: PS00674; AAA; 1.  
CC KW Cell division; ATP-binding; Transmembrane; Chloroplast.  
CC FT TRANSMEM 48 68  
CC FT TRANSMEM 96 96 POTENTIAL.  
CC FT TRANSMEM 973 993 POTENTIAL.  
CC FT TRANSMEM 1021 1041 POTENTIAL.  
CC SQ SEQUENCE 1720 AA; 197172 MW; 0F1EA926B799D5BB CRC64;

Query Match 16.9%; Score 58.5; DB 1; Length 1720;  
Best Local Similarity 26.7%; Pred. No. 57;  
Matches 16; Conservative 14; Mismatches 21; Indels 9; Gaps 4;

QY 9 VNLKDELFF-PSWEALFSGSEGLPKP-----GARIFSGDKVLRHPTW-PQKSVHWS 60  
DB 1480 IGFEIQIRYSSPMWQEDVS-AEMERKPNKPGKSLYLIVYDERTSRNBNVPPDEFYHNS 1538

RESULT 8  
FXC1\_HUMAN STANDARD; PRT; 553 AA.  
ID FXC1\_HUMAN  
AC Q12948; GUPP06;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE FORKHEAD BOX PROTEIN C1 (FORKHEAD-RELATED PROTEIN FKHL7) (FORKHEAD-  
DE RELATED TRANSCRIPTION FACTOR 3) (FREAC-3).

GN FOXC1 OR FKHL7 OR FREAC3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN (1)  
RP SEQUENCE FROM N.A., AND VARIANTS S-112; M-126 AND L-131.  
RA MEDLINE-98282091; PubMed-9620769;  
RA Nishimura D.Y., Swiderski R.E., Alward W.L.M., Searby C.C.,  
RA Patil S.R., Bennet S.R., Kanis A.B., Gastier J.M., Stone E.M.,  
RA Sheffield V.C.;  
RT "The forkhead transcription factor gene FKHL7 is responsible for  
RT glaucoma phenotypes which map to 6p25."  
RL Nat. Genet. 19:140-147(1998).  
RN (2)  
RP SEQUENCE FROM N.A., AND VARIANTS ARA THR-82 AND MET-87.  
RA MEDLINE-99011252; PubMed-9792859;  
RA Meers A.J., Jordan T., Mirzayans F., Dubois S., Kume T., Parlee M.,  
RA Ritch R., Koop B., Kuo W.-L., Collins C., Marshall J., Gould D.B.,  
RA Pearce W., Carlsson P., Enerback S., Morrisette J., Bhattacharya S.,  
RA Hogen B., Raymond V., Walter M.A.;  
RT "Mutations of the forkhead/winged-helix gene, FKHL7, in patients with  
RT Axenfeld-Rieger anomaly."  
RL Am. J. Hum. Genet. 63:1316-1328(1998).  
RN (3)  
RP SEQUENCE OF 73-178 FROM N.A.  
RA MEDLINE-95045392; PubMed-7957066;  
RA Pierron S., Helqvist M., Samuelsson L., Enerback S., Carlsson P.;  
RT "Cloning and characterization of seven human forkhead proteins:  
RT binding site specificity and DNA bending."  
RL EMO J. 13:5002-5012(1994).  
CC -1 FUNCTION: BINDING OF FREAC-3 AND FREAC-4 TO THEIR COGNATE SITES  
CC RESULTS IN BENDING OF THE DNA AT AN ANGLE OF 80-90 DEGREES.  
CC  
CC  
CC -1 SUBUNIT: MONOMER.  
CC  
CC -1 DISEASE: DEFECTS IN FKHL7 ARE THE CAUSE A SPECTRUM OF GLAUCOMA  
CC PHENOTYPES SUCH AS AXENFELD-RIEGER ANOMALY (ARA), AXENFELD-RIEGER  
CC SYNDROME (ARS) AND IRIDOGONIOSGENESIS ANOMALY (IGDA). ARS IS AN  
CC AUTOSOMAL DOMINANT DISORDER PRESENTING WITH ARA-LIKE OCULAR  
CC FINDINGS IN ADDITION TO ABNORMALITIES OF THE TEETH, JAW AND  
CC UMBILICUS.  
CC  
CC  
CC -1 SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC  
CC EMBL: AF048693; AAC18081.1; -  
CC EMBL: U13321; AAA92038.1; -  
CC EMBL: AF078096; AAC72915.1; -  
CC MIM: 601090; -  
CC MIM: 601631; -  
CC InterPro: IPR001766; -  
CC InterPro: IPR002952; -  
CC InterPro: IPR002965; -  
CC Pfam: PF00250; Fork\_head; 1.  
CC PRINTS: PR00053; FORKHEAD.  
CC PRINTS: PR01217; PRICHEXTNSN.  
CC DR PROSITE: PRO1228; EGGSHL.  
CC DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
CC DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
CC DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
CC KW DNA-binding; Nuclear protein; Transcription regulation;  
CC Disease mutation.  
CC KW DNA\_BIND 77 168 FORK-HEAD.  
CC FT DOMAIN 28 33 POLY-ALA.  
CC FT DOMAIN 169 173 POLY-ARG.  
CC FT DOMAIN 194 197 POLY-PRO.  
CC FT DOMAIN 262 272 POLY-SER.

This SWISSPORT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed, usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce#isb-sib.cn>) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

RP SEQUENCE FROM N.A.  
RC SRRATN-401:  
RX MEDLINE-88223483; PubMed-2836085;  
RA Kos T., Kuitjenhoven A., Hensing H.G.M., Pouwels P.H.,  
RA van den Hondel C.A.M.J.J.;  
RT Nucleotide sequence of the *Aspergillus niger* trpC gene: structural  
RT relationship with analogous genes of other organisms.\*;  
RL Curr. Genet. 13:137-144(1986).  
RN [2]  
RN SEQUENCE OF 1-69 AND 392-433 FROM N.A.  
RP MEDLINE-86137391; PubMed-2936650.  
RX Kos A., Kuitjenhoven J., Wernans K., Bos C.J., van den Broek H.W.J.,  
RA Pouwels P.H., van den Hondel C.A.M.J.J.;  
RT Isolation and characterization of the *Aspergillus niger* trpC gene.\*;  
RL Gene 39:231-238(1985).







```

RL Proc. Natl. Acad. Sci. U.S.A. 87:5203-5207(1990).
CC -1- FUNCTION: THIS AXONEMAL PROTEIN PARTICIPATES IN AN ENERGY SHUTTLE
CC THAT UTILIZES PHOSPHOCREATINE TO TRANSFER THE ENERGY FROM ATP
CC GENERATED BY THE MITOCHONDRION IN THE SPERM HEAD TO DYNEIN IN THE
CC DISTAL PORTIONS OF THE FLAGELLUM.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: ASSOCIATES SPECIFICALLY WITH THE AXONEME AND
CC MAY BIND DIRECTLY TO POLYMERIZED MICROTUBULES.
CC -1- DOMAIN: CONTAINS THREE COMPLETE BUT NONIDENTICAL CREATIVE KINASE
CC SEGMENTS FLANKED BY UNIQUE REGIONS.
CC -1- SIMILARITY: BELONGS TO THE ATP:GUANIDO PHOSPHOTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M33763; AAA30049.1; -.
CC PIR: A43736; A43736.
CC DR HSSP; P11009; ICRK.
CC DR InterPro; IPR000749; -.
CC DR Pfam; PF00217; ATP_gua_Ptrens; 3.
CC DR PROSITE; PS00112; GUANIDO_KINASE; 3.
CC KM Transferase; Kinase; Repeat.
CC FT REPEAT 61 414 1 (APPROXIMATE).
CC FT REPEAT 434 787 2 (APPROXIMATE).
CC FT REPEAT 808 1161 3 (APPROXIMATE).
CC FT ACT_SITE 324 324 BY SIMILARITY.
CC FT ACT_SITE 697 697 BY SIMILARITY.
CC FT ACT_SITE 1071 1071 BY SIMILARITY.
CC FT ACT_SITE 1174 AA; 130869 MW; B8F4FBA5AC48EE93 CRC64;
CC SO SEQUENCE

```

```

Query Match 16.0%; Score 55.5; DB 1; Length 1174;
Best Local Similarity 31.0%; Pred. No. 86;
Matches 13; Conservative 4; Mismatches 14; Indels 11; Gaps 2;

```

```
OY 23 LFGSGEPLKPGARIFSPGKDYLRHPTWPK-SYWHGSDPN 63
```

```
DB 607 LFDKPVSPLLTAARM-----HRDWPQGRGIMHNENKN 638
```

```

* RESULT 15
XAP5_HUMAN STANDARD; PRT; 339 AA.
ID XAP5_HUMAN
AC 014320;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE XAP-5 PROTEIN (HXC-26 PROTEIN).
GN XAP5 OR HXC26.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RP MEDLINE:97480734; PubMed:9339379;
RA Mazzarella R., Pengue G., Yoon J., Jones J., Schlessinger D.;
RT "Differential expression of XAP5, a candidate disease gene.";
RL Genomics 45:216-219(1997).
RN [2]
RP SEQUENCE OF 15-339 FROM N.A.
RP TISSUE=Skeletal muscle;
RX PubMed:9039504;
RA Toyoda A., Sakai T., Sugiyama Y., Kusuda J., Hashimoto K., Maeda H.;
RT "Isolation and analysis of a novel gene, HXC-26, adjacent to the rab
RT GDP dissociation inhibitor gene located at human chromosome Xq28
RT region.";
RL DNA Res. 3:337-340(1996).

```

```

RN [3]
RP SEQUENCE OF 66-339 FROM N.A.
RX PubMed:8733135;
RA Chen E.Y., Zolito M., Mazzarella R.A., Ciccocioppa A., Chen C.-N.,
RA Zuo L., Helner C., Burrough F.W., Ripetto M., Schlessinger D.,
RA D'Urso M.;
RT "Long-range sequence analysis in Xq28: fifteen known and six
RT candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and
RT G6PD loci.";
RL Hum. Mol. Genet. 5:659-668(1996).
CC -1- FUNCTION: MAY BE A DNA-BINDING PROTEIN OR TRANSCRIPTIONAL FACTOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED. MOSTLY
CC ABUNDANT IN FETAL BRAIN, LIVER AND KIDNEY; IN THE ADULT, HIGH
CC LEVELS WERE ALSO OBSERVED IN HEART, SKELETAL MUSCLE, SPLEEN,
CC THYMUS, PROSTATE AND SMALL INTESTINE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AD001530; AAB81663.1; -.
CC DR EMBL; D83389; BA11907.1; -.
CC DR EMBL; D83261; BA11907.1; JOINED.
CC DR EMBL; D83388; BA11907.1; JOINED.
CC DR EMBL; D83260; BA11871.1; -.
CC DR EMBL; L4140; AAA92649.1; -.
CC KM Nuclear protein.
CC FT DOMAIN 152 155 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT SEQUENCE 339 AA; 40241 MW; 88BCA57EA9B0AA32 CRC64;

```

```

Query Match 15.9%; Score 55; DB 1; Length 339;
Best Local Similarity 23.6%; Pred. No. 26;
Matches 21; Conservative 11; Mismatches 15; Indels 42; Gaps 5;

```

```
OY 8 IYNLDELIFPSWEALFS-----GSEGPLKPGARISFD----- 41
```

```
DB 249 LMYIKEDILIPHHHSFYDFIVTKARGKSGPL-----FNFVDVDDVRLLSDAFVKEKDESH 302
```

```
OY 42 -GKDYLRHPTWPK-----SYWHGSDP 62
```

```
DB 303 AGKVLR--SWYEKNKHIFPASWMEPYDP 329
```

```

Search completed: September 10, 2001, 13:50:35
Job time: 110 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2001, 13:48:05 ; Search time 13.46 Seconds  
(without alignments)  
356.538 Million cell updates/sec

Title: US-09-822-540A-1

Perfect score: 346  
Sequence: 1 ADRAAVPIVNIKDELFPSPW.....DVLRHPTPQKSWHGSDPN 63

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	346	100.0	684	2 A53019	collagen alpha 1(X
2	294	85.0	1774	2 B56101	collagen alpha 1(X
3	291	84.1	1315	2 A56101	collagen alpha 1(X
4	193	55.8	1388	2 A53317	collagen alpha 1(X
5	140	40.5	650	2 T22002	hypothetical prote
6	67.5	19.5	614	2 S27962	modulator recognit
7	64	18.5	644	2 A72519	probable 2-oxoacid
8	63.5	18.4	266	2 S22511	chlorophyll a/b-bi
9	63	18.2	336	2 T17408	rRNA (adenine-N6-)
10	62.5	18.1	467	2 T26705	hypothetical prote
11	62.5	18.1	620	2 T69797	conserved hypotet
12	62	17.9	419	2 T25565	hypothetical prote
13	61	17.6	1045	2 A29840	serine proteinase
14	60	17.3	984	2 T44496	cellulose 1,4-beta
15	59.5	17.2	268	2 H75614	hypothetical prote
16	59.5	17.2	382	2 S24434	class I histocompa
17	59	17.1	303	2 T41056	conserved hypotet
18	58.5	16.9	305	2 S07115	class I histocompa
19	58.5	16.9	354	2 I80170	class I histocompa
20	58.5	16.9	481	2 T23131	hypothetical prote
21	58.5	16.9	1184	2 D86387	probable protein p
22	58.5	16.9	1720	2 T07258	cell division prot
23	58	16.8	461	2 A54024	protein kinase (EC
24	58	16.8	662	2 T41215	probable acetate-
25	58	16.8	2055	2 T30259	multiple PDZ domai
26	57.5	16.6	506	2 H64618	sigma-54 interacti
27	57	16.5	683	1 A23690	protein kinase (EC
28	57	16.5	1228	1 JC5573	copper-transporlin
29	57	16.5	2054	2 T46612	multi PDZ domain p

30	56.5	16.3	211	2 T31724	hypothetical prote
31	56.5	16.3	217	2 T27331	hypothetical prote
32	56.5	16.3	770	2 S00643	anthranilate synth
33	56.5	16.3	884	2 C70729	hypothetical prote
34	56	16.2	651	2 F64457	methionine--C-RNA 1
35	56	16.2	753	1 D72660	probable aldehyde
36	56	16.2	838	2 T45699	hypothetical prote
37	55.5	16.0	299	2 D75138	hypothetical prote
38	55.5	16.0	553	2 T35451	ABC transporter in
39	55.5	16.0	574	2 T48113	inositol-1,4,5-tri
40	55.5	16.0	727	1 S17854	NADH dehydrogenase
41	55.5	16.0	1174	1 A43736	creatine kinase (E
42	55.5	16.0	1568	2 T08616	aggregation factor
43	55	15.9	274	2 S24439	class I histocompa
44	55	15.9	322	2 H86405	probable oxidoredu
45	55	15.9	325	2 JC5276	HXC-26 protein - h

## ALIGNMENTS

## RESULT 1

A53019  
collagen alpha 1(XVIII) chain - human (fragment)

N:Contains: endostatin

C:Species: Homo sapiens (man)

C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 31-Mar-2000

C:Accession: A53019

R:Oh, S.P.; Worman, M.L.; Seidin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olse

Genomics 19, 494-499, 1994

A:Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and local

A:Reference number: A53019; MVID:94245237

A:Accession: A53019

A:Molecule type: mRNA

A:Residues: 1-684 <OHA>

A:Cross-references: GB:122548; NID:9348908; PIDN:AA51864.1; PID:9562794

A:Note: The cited accession number, 122548, is not in Genbank release 103

A:Note: In the authors' translation, 482-Gly is not shown, residues 483-490 are shift

C:Comment: Prolines and lysines at the third position of the tripeptide repeating uni

lated and subsequently O-glycosylated.

C:Comment: different splice forms of collagen alpha 1(XVIII) may be involved in periv

C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of

ay be useful in treating solid tumors.

C:Genetics:

A:Gene: GDB:COL18A1

A:Cross-references: GDB:138752; OMIM:120328

A:Map position: 21q22.3-21q22.3

C:Superfamily: unassigned collagens

C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog

F:1-684/Product: collagen alpha 1(XVIII) chain (fragment) #status predicted <MAY>

F:1-59/Domain: collagenous (fragment) #status predicted <C05>

F:74-115/Domain: collagenous #status predicted <C04>

F:129-201/Domain: collagenous #status predicted <C06>

F:212-244/Domain: collagenous #status predicted <C07>

F:257-278/Domain: collagenous #status predicted <C08>

F:262-340/Region: cell attachment (R-G-D) motif

F:286-340/Domain: collagenous #status predicted <C09>

F:354-371/Domain: collagenous #status predicted <C010>

F:502-684/Product: endostatin #status predicted <EST>

F:509-684/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 100.0%; Score 346; DB 2; Length 684;

Best local Similarity 100.0%; Pred. No. 86-34;

Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ADRAAVPIVNIKDELFPSPWELFGSGEGPLKRGARIFSPDGRVLRHPTPQKSWHGS	60
DB	565	ADRAAVPIVNIKDELFPSPWELFGSGEGPLKRGARIFSPDGRVLRHPTPQKSWHGS	624
QY	61	DPN 63	
DB	625	DPN 627	

RESULT 2

B56101

Collagen alpha 1(XVIII) chain precursor, long splice form - mouse

M:Contains: collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin

C:Species: Mus musculus (house mouse)

C:Date: 03-Oct-1995 #sequence-revision 08-May-1998 #text-change 31-Mar-2000

C:Accession: B56101; S72450; S65595; P06675; A54072; A58816

R:Rehm, M.; Philajantemi, T.

J. Biol. Chem. 270, 4705-4711, 1995

A:Title: Identification of three N-terminal ends of type XVIII collagen chains and classification of homologous to rat and Drosophila frizzled proteins.

A:Reference number: A56101; MUID:95181468

A:Accession: B56101

A:Molecule type: mRNA

A:Residues: 1-562 <REH1>

A:Cross-references: GB:U11637; NID:9618429; PIDN:AAC52179.1; PID:9618430

A:Experimental source: splice form clone PE17.24

A:Accession: C56101

A:Molecule type: mRNA

A:Residues: 1-239, 487-562 <REH2>

A:Cross-references: GB:U11637; NID:9618429

A:Experimental source: splice form clones PE8.1, PE19, PE15.2

R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.

submitted to the EMBL Data Library, August 1993

A:Reference number: S72450

A:Accession: S72450

A:Molecule type: mRNA

A:Residues: 487-1146, 'L', 1148-1193, 'F', 1195-1210, 'R', 1212-1512, 'L', 1514-1522, 'F', 1524-16

A:Cross-references: EMBL:L22545; NID:9348968; PIDN:AA19787.1; PID:9511298

R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.

Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994

A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa

A:Reference number: A58370; MUID:94240111

A:Accession: S65595

A:Molecule type: mRNA

A:Residues: 487-1512, 'L', 1514-1522, 'F', 1524-1683, 'V', 1685-1774 <OH2>

A:Cross-references: EMBL:L22545

R:Abel, N.; Muragaki, Y.; Yoshioke, H.; Inoue, H.; Nishimura, Y.

Biochem. Biophys. Res. Commun. 196, 576-582, 1993

A:Title: Identification of a novel collagen chain represented by extensive interruptions

A:Reference number: P06675; MUID:94059075

A:Accession: P06675

A:Molecule type: mRNA

A:Residues: 635-1774 <ABD>

R:Rehm, M.; Hantelike, E.; Philajantemi, T.

J. Biol. Chem. 269, 13929-13935, 1994

A:Title: Primary structure of the alpha chain of mouse type XVIII collagen, partial structure

A:Reference number: A54072; MUID:94245707

A:Accession: A54072

A:Molecule type: DNA

A:Residues: 1293-1403, 'R', 1405-1774 <REH3>

A:Cross-references: GB:U03714; NID:9487733; PIDN:AAA20657.1; PID:9487734

R:O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukui, N.; Vasios, G.; Lane, W.S.; Flynn, E.; Bl

Cell 88, 277-285, 1997

A:Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.

A:Reference number: A58816; MUID:97160848

A:Accession: A58816

A:Molecule type: protein

A:Residues: 1591-1610 <ORE>

A:Experimental source: hemangioendothelium cells

A:Note: Inhibits endothelial cell proliferation

C:Comment: Prolins and lysines at the third position of the tripeptide repeating unit labeled and subsequently O-glycosylated

C:Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in part

C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un

ay be useful in treating solid tumors.

C:Genetics:

A:Gene: MGI:COL18a1

A:Cross-references: MGI:71175

A:Map position: 10:41.0

A:Introns: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 15

A:Note: the list of introns is incomplete

C:Superfamily: unassigned collagens

C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog

F:1-1774/Product: collagen alpha 1(XVIII) chain precursor, long splice form #status p

F:1-239,487-1774/Product: collagen alpha 1(XVIII) chain precursor, medium splice form

F:1-24/Domain: signal sequence #status predicted <SIG>

F:361-486/Region: frizzled similarity

F:786-812/Domain: collagenous #status predicted <CO01>

F:823-896/Domain: collagenous #status predicted <CO02>

F:921-1042/Domain: collagenous #status predicted <CO03>

F:1066-1148/Domain: collagenous #status predicted <CO04>

F:1163-1304/Domain: collagenous #status predicted <CO05>

F:1301-1333/Domain: collagenous #status predicted <CO06>

F:1346-1369/Domain: collagenous #status predicted <CO07>

F:1351-1353/Region: cell attachment (R-G-D) motif

F:1377-1428/Domain: collagenous #status predicted <CO08>

F:1442-1459/Domain: collagenous #status predicted <CO09>

F:1591-1774/Product: endostatin #status predicted <EST>

F:1598-1774/Region: multiplexin collagen carboxyl-terminal similarity

F:354,361,947/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted

F:910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 85.0%; Score 294; DB 2; Length 1774;

Best Local Similarity 81.0%; Pred. No. 4,7e-27;

Matches 51; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADRAVPYINLDELFPSEALFSGSECPKRGARIFSPDCKDVLRRPMPKQSWHGS 60

DB 1654 ADRCVSPIYNLKDVELSPRWSLPSGSGQQLDPGARIFSPDGRDVLRRPMPKQSWHGS 1713

QY 61 DPN 63

DB 1714 DPS 1716

RESULT 3

A56101

Collagen alpha 1(XVIII) chain precursor, short splice form - mouse

M:Contains: endostatin

C:Species: Mus musculus (house mouse)

C:Date: 03-Oct-1995 #sequence-revision 08-May-1998 #text-change 31-Mar-2000

C:Accession: A56101; A58371; S72450; S65595

R:Rehm, M.; Philajantemi, T.

J. Biol. Chem. 270, 4705-4711, 1995

A:Title: Identification of three N-terminal ends of type XVIII collagen chains and classification homologous to rat and Drosophila frizzled proteins.

A:Reference number: A56101; MUID:95181468

A:Accession: A56101

A:Molecule type: mRNA

A:Residues: 1-103 <REH1>

A:Cross-references: GB:U11636; NID:9618427; PIDN:AAC52178.1; PID:9618428

R:Rehm, M.; Philajantemi, T.

Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994

A:Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collageno

A:Reference number: A58371; MUID:94240112

A:Accession: A58371

A:Molecule type: mRNA

A:Residues: 1-928 <REH2>

A:Cross-references: GB:U16898; NID:9404754; PIDN:AAA37434.1; PID:9553894

R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.

submitted to the EMBL Data Library, August 1993

A:Reference number: S72450

A:Accession: S72450

A:Molecule type: mRNA

A:Residues: 28-687, 'L', 689-734, 'F', 736-751, 'R', 753-1315 <OHW>

A:Cross-references: EMBL:L22545; NID:9348968; PIDN:AA19787.1; PID:9511298

R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.

Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994

A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-

[illegible]

```

A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-9,'S','11-48','V','50-94','A','96-149','A','151-203','V','205-408','A','410-569' <M>
A:Cross-references: GB:D21230; MID:9415605; PIDN:BA04762.1; PID:d1005294; PID:g46070
R:Meers,J.C.; Kivirikko,S.; Gordon,M.R.; Dion,A.S.; Pihlajaniemi,T.
Proc. Natl. Acad. Sci. U.S.A. 89, 10144-10148, 1992
A>Title: Identification of a previously unknown human collagen chain, alpha1(XV), cha
A:Reference number: S28778; MUID:93066196
A:Accession: S28778
A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 544-640,'P','642-811','P','813-1252' <MYE>
C:Genetics:
A:Gene: GDB:COL15A1
A:Cross-references: GDB:132578; OMIM:120325
A:Map position: 9q21-q9q22
C:Superfamily: unassigned collagens
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-1388/Product: collagen alpha 1(XV) chain #status predicted <MAT>
F:1216-1388/Region: multiplexin collagen carboxyl-terminal similarity

Query Match          55.8%; Score 193; DB 2; Length 1388;
Best Local Similarity 49.2%; Pred. No. 5,7e-15;
Matches 31; Conservative 16; Mismatches 16; Indels 0; Gaps 0;

OY      1 ADRAAPVTNLKDELLFSGSEGPLKPGARIFSPDGKDYLRHPPTQKSWMHGS 60
       1:::|||||:::|:::|::| 1 1::|::|::| 1::|::|::|
Db      1268 AERYSLPTVLNKGYLFWNMDSIFGSHGGOFMMHPIPYISFDGRIDMTDPSPQKIWMHGS 1327

OY      61 DPN 63
       1:
Db      1328 SPH 1330

RESULT   5
T22002
hypothetical protein F39H11.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22002
R:White, S.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19500
A:Accession: T22002
A>Status: Preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-650 <MIT>
A:Cross-references: EMBL:Z81079; PIDN:CAB03084.1; GSPDB:GN00019; CESP:F39H11.4
A:Experimental source: clone F39H11
C:Genetics:
A:Gene: CESP:F39H11.4
A:Map position: 1
A:Introns: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

Query Match          40.5%; Score 140; DB 2; Length 650;
Best Local Similarity 47.2%; Pred. No. 5,9e-09;
Matches 25; Conservative 9; Mismatches 17; Indels 2; Gaps 1;

OY      8 IYNLKDELLFSGSEGPLKPGARISFSDGKDYLRHPPTQKSWMHGS 60
       1::|::|::|::|::|::| 1::|::|::| 11 1 11111
Db      525 VNVAGHHLPMSRSFVNQAQ--MNPNAKLFFSDRHVDLVDSRWPDKRWMHGS 575

RESULT   6
S27962
modulator recognition factor 1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 05-Nov-1999
C:Accession: S27962
R:Okada, T.; Merrills, B.W.; Huang, E.; Okada, T.; Tanaka, Y.; Gertson, P.; Itakura, K.
Submitted to the EMBL Data Library, March 1991

```

A:Reference number: S27962  
A:Accession: S27962  
A:Molecule type: mRNA  
A:Residues: 1-614 <OKA>  
A:Cross-references: EMBL:MG324; NID:g188683; PIDN:AAA36325.1; PID:g188684;

Query Match	19.5%	Score 67.5;	DB 2;	length 614;
Best Local Similarity	32.8%;	Pred. NO. 3.2;		
Matches 20; Conservative	7;	Mismatches 23;	Indels 11;	Gaps 3;

```

0Y      3 RAAPVIVNLKDELLFFSWEALFSGSEGLKPGARIF-----SFDCKDVLKRHPWPOKSW 57
      ::::  |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
Db      541 QAALPF----SPLVITPAFPAHFLATAGTSPMAAGIMHPPTSFD--SALRRRLCPASSAW 594

```

QY	58	H	58
		1	
Db	595	H	595

\*RESULT 7  
A72519

C:Accession: A72518  
C:Date: 20-Aug-1999  
C:Species: Aeropyrum pernix  
P:Probable 2-Oxoglutarate-dependent oxidoreductase alpha chain APE3126 - Aeropyrum pernix (strain ATCC 49239)  
#sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000

R:Kikarabaeysai,Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hatakeyama, Y.; Ueno, K.; Takahashi, T.; Tanaka, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Nishida, A.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Nishida, A. Res. 6: 83-101, 1999

A>Title: Complete genome sequence of an aerobic hyper-thermophilic Gram-negative bacterium, *Aeropyrum pernix*

X:Reference number: A72450; MUID:99310339

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-644 <KAM>  
A:Cross-references: DDBJ:AP000063; NID:g5105654; PIDN:BAAB137.1; PID:g5105825  
A:Experimental source: strain K1  
A:Genetics:

A; Gene: APE2126  
C; Superfamily: Halobacterium halobium 2-oxoacid ferredoxin oxidoreductase; 2-oxoacid ferredoxin

Query Match	18.5%	Score 64;	DB 2;	Length 644;
Best Local Similarity	26.6%;	Pred. NO. 9;		
Matches 21; Conservative	9;	Mismatches 29;	Indels 20;	Gaps 3

```

OY      1  ADRAVPVIVLNKDELL-----PSPMEA-----LFSGSEGPLKPCARIFSEFGKD 44
         1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db      408 AERQPMVTHLLDKLANMVASVPFRPMKAIIDRGKTLFKAPRGPFKRPDPDPLDRP 467

```

```
QY      45 VLRIPTPQKSVNHGSDPN 63
          ||      | : | |
Db      468 VLGSGA---ITWYTGDEN 482
```

RESULT 8  
S22511  
chlorophyll a/b-binding protein precursor - white mustard  
C:Species: Sinapis alba (white mustard)  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 26-Aug-1999  
C:Accession: S22511: S06765  
R:Gully, A.; Batschauer, A.; von Arnim, A.; Koesel, H.  
Plant Mol. Biol. 19, 277-287, 1992  
O:Title: isolation and characterization of a gene encoding a chlorophyll a/b-binding protein

A/Reference number: S22511; MUID:923229555  
A/Accession: S22511

A; molecule type: DNA

A;ResIdues: 1-266 <GAU>  
A;CROSS-references: FMP

n/c1050 sciences. EMBL:AL0430, MID:g21137, P1DN:CAA34439.1, P1D:g21138  
 C:Gorot100:

A;Gene: cab-1

C<sub>2</sub>Superfamily: chlorophyll a/b-binding protein

**C**:Keywords: chlorophyll; chloroplast; light-harvesting complex; Membrane protein; thylakoid; F-1-3/Domain: transit peptide (Chloroplast) #status predicted <TN>  
F:35-266/Product: chlorophyll a/b-binding protein #status predicted <MT>

Query Match	18.48;	Score 63.5;	DB 2;	Length 266;
Best Local Similarity	39.08;	Pred. No. 3.6;		
Matches 16;	Conservative 4;	Mismatches 18;	Indels 3;	Gaps 1

```

QY 24 FSGSEGPLKPGARIFSFDDGKDLR---HP*PWQKSVHHGSD 61
      | | | | | : | | | | | | | | |
Db 13 FAGKAVKLSFGASEVFCTGRVTMRKTYKPPGSGSPHWGSD 53

```

RESULT 9  
T17408

rRNA (adenine-N6)-methyltransferase (EC 2.1.1.48) 1 - Streptomyces venezuelae  
 C:Species: Streptomyces venezuelae  
 C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 08-Sep-2000  
 C:Accession: M17408

R:Xue, Y.; Zhao L.; Liu, H.W.; Sherman, D.H.  
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998  
A.Titles: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae

A:Accession Number: Z10773, MOLID:5844333  
A:Accession: 117408  
A:Status: preliminary; translated from GB/EMBL/DBJ  
Molecule type: DNA

A:Residues: 1-330 <XUES>  
A:Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800833; PIDN:MAC69328.1  
C:Gene: pikr1  
A:Gene: pikr1

**C;Keywords:** methyltransferase; S-adenosylmethionine

Query Match	18.2%	Score 63;	DB 2;	Length 336;
Best Local Similarity	40.0%;	Pred. NO. 5.5;		
Matches	14;	Conservative	3;	Mismatches 18;
			Indels	0;
			Gaps	0;

Oy	1	ADRAAVPIVNKDELLEFP	SWEALFSSGSEGPLRGA	35
Db	235	ADRHGIPDGGLEPKDLTTQWIALEQASQPSYAPGA	269	

RESULT 10  
T26705  
hypothetical protein Y38H6C.17 - *Caenorhabditis elegans*

Species: *Caenothabax elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T26705  
R:White, S.

A:Reference number: 220255  
 submitted to the EMBL Data Library, September 1998  
 A:Accession: T26705  
 S:Status: Preliminary: translated from GB/EMBL/DBP

A: Molecule type: DNA  
A: Residues: 1-467 <MIL>  
A: Cross-references: EMBL:AL031630; PDB:CAA20995.1; GSPDB:GN00023; CESP:Y38H6C.17  
Experimental source: clone Y38H6C

C/Genetics:  
A;Gene: CESP:Y38H6C.17  
A;Map position: 5  
A;Intons: 9/2: 49/2: 79/3: 109/2: 140/2: 168/3: 225/3: 322/2: 382/1: 429/2:

Query Match	18.1%;	Score 62.5;	DB 2;	Length 467;
Best Local Similarity	33.9%;	Pred. No. 9.4;		
Matches	20;	Conservative	12;	Mismatches 15;
				Gaps 5;

QY 4 AAMPVNLKDELL-----FPSWE-ALFSGSEG-PLKPGARISFDG-----KDVLRHP 49  
 218 AATALT--QELSLHWYPTWELPSITGEGVSLAGSLITISFEGQANVLPLENSLKH 274

## RESULT 11

F69797

conserved hypothetical protein yeww - *Bacillus subtilis*  
C:Species: *Bacillus subtilis*

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000

C:Accession: F69797

R:Kunst, F.; Ogata, N.; Mosser, I.; Albertini, A.M.; Alloul, G.; Azevedo, V.; Berlet, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Hensut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois, A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y.M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Paro, V.; Pohl, T.M.; Portelle, Rieger, M.; Rivalet, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schoeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K. A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. A:Reference number: A69580; MUID:98044033

A:Accession: F69797

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-620 &lt;KUN&gt;

A:Cross-references: GB:299107; GB:AL009126; NID:92632866; PIDN:CAB12524.1; PID:92633018

A:Experimental source: strain 168

C:Genetics:

A:Gene: yeww

## Query Match

Best Local Similarity 18.1%; Score 62.5; DB 2; Length 620;  
Pred. No. 13;

Matches 23; Conservative 10; Mismatches 16; Indels 49; Gaps 4;

QY 12 KDELFPSP-----EALFSGSEGLPKGARIFS----- 39

DB 375 KDELIFGSMVADHDGKMGSTGIGMDALHTGDLDPGRGLFVQYHEDKNAYGLSFRD 434

QY 40 -----FDGKDYLR-----HPTWPKSYW-HGS 60

DB 435 AATGKLMGVYAGKVDGKGMADIDPRYQGEYMWANGS 472

## RESULT 12

T25565

hypothetical protein C24A8.1 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000

C:Accession: T25565

R:Nelson, J.; Gilling, S. submitted to the EMBL Data Library, December 1996

A:Description: The sequence of *C. elegans* cosmid C24A8.

A:Reference number: Z20051

A:Accession: T25565

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-419 &lt;NEU&gt;

A:Cross-references: EMBL:U80845; PIDN:AAB37978.1; GSPDB:GN00028; CESP:C24A8.1

A:Experimental source: strain Bristol N2; clone C24A8

C:Genetics:

A:Gene: CESP:C24A8.1

A:Map position: X

A:Introns: 27/3; 49/3; 87/2; 186/3; 213/1; 250/3; 283/1; 348/3; 374/2; 407/1

C:Superfamily: vertebrate rhodopsin

## Query Match

Best Local Similarity 17.9%; Score 62; DB 2; Length 419;  
Pred. No. 9.5;

Matches 16; Conservative 4; Mismatches 17; Indels 0; Gaps 0;

QY 4 AAVPYVNLKDELFPSPWEALFSGSEGLPKGARIFSF 40

DB 8 AAVAMELEDELQPEMAYLVSLAMADLIVGAFVMPF 44

## RESULT 13

A29840

serine proteinase (EC 3.4.21.-) precursor - *Serratia marcescens* (strain IFO-3046)

C:Species: *Serratia marcescens*

C:Date: 19-May-1989 #sequence\_revision 19-May-1989 #text\_change 22-Jun-1999

C:Accession: A29840

R:Yanagida, N.; Uozumi, T.; Beppu, T.

J. Bacteriol. 166, 937-944, 1986

A:Title: Specific excretion of *Serratia marcescens* protease through the outer membrane

A:Reference number: A29840; MUID:86223815

A:Accession: A29840

A:Molecule type: DNA

A:Residues: 1-1045 &lt;YAN&gt;

A:Cross-references: GB:M13469; NID:9152857; PIDN:AAA26572.1; PID:9152858

C:Superfamily: Serratia serine proteinase; subtilisin homology

C:Keywords: hydrolase; serine proteinase

F:1-27/Domain: signal sequence #status predicted <SIG>

F:28-408/Product: serine proteinase #status predicted <MAT>

F:67-355/Domain: subtilisin homology <SPR>

F:409-1045/Domain: carboxyl-terminal propeptide #status predicted <PRO>

F:76-112,341/Active site: Asp, His, Ser #status predicted

## Query Match

Best Local Similarity 17.6%; Score 61; DB 2; Length 1045;  
Pred. No. 37;

Matches 15; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 14 ELFPSPWEALFSGSEGLPKGARIFSGDKVDLRHPTWPKSY 56

DB 449 ECFSDSMNDISGHGLTKGACTLALLGNNTYRGDTWYKQGV 491

## RESULT 14

T44496

cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) precursor [imported] - *Thermomonospora*

N:Alternate names: beta-1,4-exocellulase E6 precursor

C:Species: *Thermomonospora fusca*

C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jan-2000

C:Accession: T44496

R:Irwin, D.I.; Zhang, S.; Wilson, D.B.

submitted to the EMBL Data Library, April 1999

A:Description: Characterization of a *Thermomonospora fusca* family 48 exocellulase E6.

A:Reference number: Z22783

A:Accession: T44496

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-984 &lt;IRW&gt;

A:Cross-references: EMBL:AF144563; PIDN:AA39947.1

A:Experimental source: strain YX

C:Genetics:

A:Gene: celf

C:Keywords: glycosidase; hydrolase

F:1-34/Domain: signal sequence #status predicted <SIG>

F:35-984/Product: cellulose 1,4-beta-cellobiosidase #status predicted <MAT>

Query Match 17.3%; Score 60; DB 2; Length 984;  
Best Local Similarity 33.3%; Pred. No. 45;

Matches 22; Conservative 5; Mismatches 13; Indels 26; Gaps 5;

QY 18 PSWEAL---FSGSEG-----PLKPGA---RISFDGKVDLRHPTWPKSY 56

DB 889 PSWDRDDPMDGSEGLYVPGWGTMPNGDRIDPGATFLSRFYRND---PLMPQVFA 944

QY 57 WHGSDP 62

DB 945 -HLNDP 949

RESULT 15  
H75614  
hypotheical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: H75614  
R:White, O.; Elsen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: H75614  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-268 <WHI>  
A:Cross-references: GB:AE001862; GB:AE001825; NID:96460468; PIDN:AAFI2307.1; PID:9646060  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DRA0182  
A:Map position: 2

Query Match 17.28; Score 59.5; DB 2; Length 268;  
Best Local Similarity 36.1%; Pred. No. 11;  
Matches 13; Conservative 7; Mismatches 15; Indels 1; Gaps 1;  
Oy . 13 DELFPSSWALFSGSGPL-KPGARIFSPDGKDLR 47  
| : : : | : : | | | | : | | |  
Db 165 DDASTAALDRLLAQAGPLSNPAGRIFFPENNDALR 200

Search completed: September 10, 2001, 13:49:50  
Job time: 105 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: September 10, 2001, 13:47:45 : Search time 12.18 Seconds

(without alignments)  
106.502 Million cell updates/sec

Title: US-09-822-540A-1

Perfect score: 346  
Sequence: 1 ADRAAVPIVNLKDELFPSW.....DYLHPTWPKSVWHS DPN 63Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0  
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%  
Listing first 45 summaries

## Database :

Issued Patents-AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/6CTUS.COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	346	100.0	183	4	US-09-206-059-2
2	291	84.1	195	1	US-08-159-784-2
3	277.5	80.2	185	3	US-08-985-526-36
4	190	54.9	191	1	US-08-159-784-3
5	94	27.2	16	4	US-09-385-442-32
6	59	17.1	14	4	US-09-385-442-31
7	57.5	16.6	506	2	US-08-849-480A-5
8	57.5	16.6	553	2	US-09-083-351-2
9	57.5	16.6	553	4	US-09-083-352-2
10	55	15.9	444	1	US-08-483-140-28
11	55	15.9	444	2	US-08-485-938A-32
12	54.5	15.8	470	2	US-08-377-440A-1
13	54.5	15.8	578	1	US-08-653-740-3
14	54.5	15.8	578	2	US-09-073-594-3
15	54.5	15.8	578	3	US-09-275-925-3
16	54.5	15.8	636	1	US-08-653-740-5
17	54.5	15.8	636	2	US-09-073-594-5
18	54.5	15.8	636	3	US-09-275-925-5
19	53	15.3	312	4	US-09-216-295-21
20	53	15.3	371	4	US-09-104-308-1
21	52.5	15.2	293	1	US-08-628-291-4
22	52.5	15.2	293	2	US-09-128-722-4
23	52.5	15.2	317	1	US-08-628-291-12
24	52.5	15.2	317	2	US-09-128-722-12
25	52.5	15.2	372	1	US-08-149-093A-6
26	52.5	15.2	372	2	US-08-911-245-6
27	52.5	15.2	372	2	US-08-514-451A-9

28	52.5	15.2	372	2	US-08-411-859-2	Sequence 2, Appl1
29	52.5	15.2	372	2	US-08-411-859-10	Sequence 10, Appl1
30	52.5	15.2	372	3	US-09-170-331-6	Sequence 6, Appl1
31	52.5	15.2	372	3	US-08-147-592A-4	Sequence 4, Appl1
32	52.5	15.2	372	4	US-08-430-266A-6	Sequence 6, Appl1
33	52	15.0	449	3	US-08-987-743-7	Sequence 7, Appl1
34	52	15.0	514	4	US-07-796-899-25	Sequence 25, Appl1
35	52	15.0	841	4	US-09-413-814-107	Sequence 107, Appl1
36	51	14.7	444	2	US-08-485-938A-33	Sequence 33, Appl1
37	50.5	14.6	164	2	US-08-918-723-3	Sequence 3, Appl1
38	50.5	14.6	164	2	US-09-237-507-3	Sequence 2, Appl1
39	50.5	14.6	294	1	US-08-628-291-2	Sequence 2, Appl1
40	50.5	14.6	294	1	US-09-128-722-2	Sequence 2, Appl1
41	50.5	14.6	318	1	US-08-628-291-16	Sequence 16, Appl1
42	50.5	14.6	318	2	US-09-128-722-16	Sequence 16, Appl1
43	50.5	14.6	1235	2	US-08-680-326-36	Sequence 36, Appl1
44	50	14.5	31	3	US-09-045-632-77	Sequence 77, Appl1
45	50	14.5	98	3	US-09-045-632-5	Sequence 5, Appl1

## ALIGNMENTS

RESULT 1  
US-09-206-059-2  
; Sequence 2, Application US/09206059  
; Patent No. 6201104  
GENERAL INFORMATION:  
; APPLICANT: MacDonald, Nicholas  
; APPLICANT: Sim, Kim Lee  
; TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and  
; FILE REFERENCE: 05213-0370  
; CURRENT APPLICATION NUMBER: US/09/206,059  
; CURRENT FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 183  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-206-059-2

Query Match 100.0%; Score 346; DB 4; Length 183;  
Best Local Similarity 100.0%; Pred. No. 2e-38;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELFPSWEALFSGSEGPLKPGARIFSDGKDYLRHPTWPKSVWHS 60  
DB 64 ADRAAVPIVNLKDELFPSWEALFSGSEGPLKPGARIFSDGKDYLRHPTWPKSVWHS 123

QY 61 DPN 63  
DB 124 DPN 126

RESULT 2  
US-08-159-784-2  
; Sequence 2, Application US/08159784  
; Patent No. 5643783  
GENERAL INFORMATION:  
; APPLICANT: Bjorn R. Olsen  
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF  
; NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 555X  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/159,784  
FILING DATE: December 1, 1993  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: John F. Freeman  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 00246/170001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 195  
TYPE: amino acid  
STRANDEDNESS: N/A  
TOPOLOGY: N/A  
US-08-159-784-2

Query Match 84.1% Score 291; DB 1; Length 195;  
Best Local Similarity 79.4%; Pred. No. 4, 1e-31;  
Matches 50; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADRAAPVIVNLKDELFPSEALFSGSGEPLKPGARIFSPDGKDLRHPTWPKSWHGS 60  
DB 75 ADGSGVPIVNLKDELFPSEALFSGSGEPLKPGARIFSPDGKDLRHPTWPKSWHGS 134

QY 61 DPN 63  
DB 135 DPS 137

RESULT 3  
US-08-985-526-36  
Sequence 36, Application US/08985526  
Patent No. 6080728  
GENERAL INFORMATION:  
APPLICANT: Mixson, James A  
TITLE OF INVENTION: CARRIER-DNA COMPLEXES CONTAINING DNA  
TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE  
TITLE OF INVENTION: THERAPY  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Connolly, Boyle, Lodge, & Hutz  
STREET: 1220 Market Street, P.O. Box 2207  
CITY: Wilmington  
STATE: Delaware  
COUNTRY: U.S.A.  
ZIP: 19899  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/985,526  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/608,845  
FILING DATE: 16-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: McMorow Jr., Robert G  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (302) 658-9141  
TELEFAX: (302) 658-5613  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 185 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-985-526-36

Query Match 80.2% Score 277.5; DB 3; Length 185;  
Best Local Similarity 76.6%; Pred. No. 2, 4e-29;  
Matches 49; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

QY 1 ADRAAPVIVNLKDELFPSEALFSGSGEPLKPGARIFSPDGKDLRHPTWPKSWHGS 59  
DB 64 ADGSGVPIVNLKDELFPSEALFSGSGEPLKPGARIFSPDGKDLRHPTWPKSWHGS 123

QY 60 SDPN 63  
DB 124 SDPS 127

RESULT 4  
US-08-159-784-3  
Sequence 3, Application US/08159784  
Patent No. 5643783  
GENERAL INFORMATION:  
APPLICANT: Bjorn R. Olsen  
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 555X  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/159,784  
FILING DATE: December 1, 1993  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: John F. Freeman  
REGISTRATION NUMBER: 29,066  
REFERENCE/DOCKET NUMBER: 00246/170001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 191  
TYPE: amino acid  
STRANDEDNESS: N/A  
TOPOLOGY: N/A  
US-08-159-784-3

Query Match 54.9% Score 190; DB 1; Length 191;  
Best Local Similarity 49.2%; Pred. No. 9, 9e-18;  
Matches 31; Conservative 16; Mismatches 16; Indels 0; Gaps 0;  
QY 1 ADRAAPVIVNLKDELFPSEALFSGSGEPLKPGARIFSPDGKDLRHPTWPKSWHGS 60  
DB 75 ADGSGVPIVNLKDELFPSEALFSGSGEPLKPGARIFSPDGKDLRHPTWPKSWHGS 134



Query Match	16.6%	Score 57.5	DB 4	length 553
Best Local Similarity	37.9%	Pred. No. 14		
Matches	22	Conservative	2	Mismatches 29
				Indels 5
				Gaps 2
Oy	4	AAAPVIVNLKDELLFPSEWALFSGSECP-LKPCARLRFSGDKVYLRIHPVTPQKSWHGS	60	
Db	252	AAAPVIESPDLS---SSSSLSGSSSPGSLPFAKRLSLGAGASAPPPAPSPAPPHNS	305	

RESULT 9  
US-09-083-352-2  
Sequence 2, Application US/09083352  
Patent No. 6207450  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Darryl  
APPLICANT: Pacil, Shiva  
TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS  
TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-483-140-28

Query Match 15.9%; Score 55; DB 1; Length 444;  
Best Local Similarity 30.8%; Pred. No. 23;  
Matches 12; Conservative 7; Mismatches 20; Indels 0; Gaps 0;

Qy 17 FPSWEALFSGSGPLKPGARIFSPDGKDVLRHPTWPKS 55  
Db 403 PDQMSLIEGKDENLMPGTNITNEHTLONSPEAKS 441

RESULT 11  
US-08-485-938A-32  
; Sequence 32, Application US/08485938A  
; Patent No. 5847088  
; GENERAL INFORMATION:  
; APPLICANT: Cousens, Lawrence S.  
; APPLICANT: Eberhardt, Christine D.  
; APPLICANT: Gray, Patrick W.  
; APPLICANT: Le Trong, Hai  
; APPLICANT: Tjoelker, Larry W.  
; APPLICANT: Wilder, Cheryl L.  
; TITLE OF INVENTION: Platelet-Activating Factor  
; TITLE OF INVENTION: Acetylhydrolase  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,938A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/318,905  
; FILING DATE: 06-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/133,803  
; FILING DATE: 06-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5847088and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 27866/32792  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 474-6300  
; TELEFAX: (312) 474-0448  
; TELEX: 25-3658  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 444 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-485-938A-32

Query Match 15.9%; Score 55; DB 2; Length 444;  
Best Local Similarity 30.8%; Pred. No. 23;  
Matches 12; Conservative 7; Mismatches 20; Indels 0; Gaps 0;  
Qy 17 FPSWEALFSGSGPLKPGARIFSPDGKDVLRHPTWPKS 55

Db 403 PDQMSLIEGKDENLMPGTNITNEHTLONSPEAKS 441

RESULT 12  
US-08-377-440A-1  
; Sequence 1, Application US/08377440A  
; Patent No. 5985623  
; GENERAL INFORMATION:  
; APPLICANT: POLLOCK, THOMAS J.  
; APPLICANT: YAMAZAKI, MOTOHIDE  
; APPLICANT: THORNE, LINDA  
; APPLICANT: MIKOLAJCZAK, MARCIA  
; APPLICANT: ARMENTROUT, RICHARD W.  
; TITLE OF INVENTION: DNA SEGMENTS AND METHODS FOR INCREASING  
; TITLE OF INVENTION: POLYSACCHARIDE PRODUCTION  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: JULES E. GOLDBERG  
; STREET: 261 MADISON AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10016-2391  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/377,440A  
; FILING DATE: 24-JAN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GOLDBERG, JULES E.  
; REGISTRATION NUMBER: 24,408  
; REFERENCE/DOCKET NUMBER: JG-SEC-1166  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-986-4090  
; TELEFAX: 212-818-9479  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 470 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-08-377-440A-1

Query Match 15.8%; Score 54.5; DB 2; Length 470;  
Best Local Similarity 36.8%; Pred. No. 28;  
Matches 14; Conservative 8; Mismatches 7; Indels 9; Gaps 2;

Qy 9 VNIKDELFPSWEALFSGSGPLKPGARIFSPDGKDVLR 46  
Db 234 NNIKEILVPPFNAL--GAIG-----VDSYEGKDTL 262

RESULT 13  
US-08-653-740-3  
; Sequence 3, Application US/08653740  
; Patent No. 5792850  
; GENERAL INFORMATION:  
; APPLICANT: James W. Baumgartner  
; APPLICANT: Donald C. Foster  
; APPLICANT: Frank J. Grant  
; APPLICANT: Cindy A. Sprecher  
; TITLE OF INVENTION: HEMATOPOIETIC CYTOKINE RECEPTOR  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Zymogenetics, Inc.  
; STREET: 1201 Eastlake Avenue East



Db	85	AGRSWVAIPREQLTMSDKLLVWGCTKAGQPLMPPVFVNLETOMKPNAPRLGPPDVDFSEDDP	144
Oy	46	LR-----HPTWPOKSV	56
Db	145	LEATVHWAPPTWPSHKV	161

Search completed: September 10, 2001, 13:49:29  
Job time: 104 sec

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2001, 13:47:25 ; Search time 20.16 Seconds

(without alignments)  
189,450 Million cell updates/sec

Title: US-09-822-540A-1

Perfect score: 346  
Sequence: 1 ADRAAVPTVNLKDELFPW.....DVLHPTWPKSVWNGSDPN 63

Scoring table:  
BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database :

A\_Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	346	100.0	178	21	AAV94324
2	346	100.0	182	21	AAAB28399
3	346	100.0	182	21	AAV94323
4	346	100.0	182	21	AAV569622
5	346	100.0	183	20	AAV086993
6	346	100.0	183	20	AAV02113
7	346	100.0	183	21	AAAB30493
8	346	100.0	183	21	AAAB16451
9	346	100.0	183	21	AAV90771
10	346	100.0	183	21	AAV70252
11	346	100.0	183	22	AAAB9379

12	346	100.0	193	21	AAW90877	Human HMW endostat
13	346	100.0	195	21	AAW90874	Human HMW endostat
14	346	100.0	216	21	AAAB30495	Amino acid sequenc
15	346	100.0	684	18	AAW26327	Human alpha-1 coll
16	346	100.0	684	20	AAV25113	Human alpha-1 (XVI
17	346	100.0	1301	20	AAW92296	Human collagen 18
18	346	100.0	1336	20	AAV086594	Human endostatin p
19	342	98.8	183	22	AAAB49810	A human angiogenes
20	335	96.8	271	21	AAAB08407	Canine angiogenesi
21	302	87.3	184	21	AAV70265	Murine endostatin
22	284	85.0	108	22	AAAB9667	Murine endostatin
23	284	85.0	184	22	AAAB49380	Murine endostatin
24	294	85.0	191	21	AAAB28398	Murine endostatin
25	293	84.7	184	20	AAV18409	Endostatin protein
26	291	84.1	184	20	AAV08689	Murine endostatin
27	291	84.1	184	21	AAV70258	Murine angiogenesi
28	281	84.1	207	22	AAAB71830	Murine endostatin
29	281	84.1	218	20	AAV08691	Murine gene therap
30	291	84.1	580	20	AAV08692	Murine gene therap
31	291	84.1	684	20	AAV25114	Mouse alpha1 (XVII
32	291	84.1	1288	18	AAW26328	Mouse alpha-1 coll
33	291	84.1	1288	20	AAW92297	Mouse alpha-1 (XVI
34	277.5	80.2	185	20	AAV06197	Anti-angiogenic en
35	255	73.7	184	22	AAAB49381	Chicken endostatin
36	197	56.9	180	22	AAAB49383	Murine endostatin
37	193	55.8	176	21	AAV90875	Human HMW endostat
38	193	55.8	180	22	AAAB49382	Human endostatin S
39	193	55.8	181	20	AAV22227	Human reslin prote
40	193	55.8	181	20	AAV18410	Reslin protein seq
41	193	55.8	191	21	AAAB90876	Human HMW endostat
42	182	52.6	45	22	AAAB35586	Antiangiogenic pen
43	114	32.9	85	20	AAV22226	Apomigren. Homo s
44	109	31.5	50	22	AAAB35585	Antiangiogenic pen
45	107	30.9	20	22	AAAB49856	Human endostatin p

#### ALIGNMENTS

RESULT 1  
ID AAY94324 standard; Protein; 178 AA.  
AC AAY94324;  
XX  
XX  
DT 11-AUG-2000 (first entry)  
XX  
XX  
DE Alternate human endostatin protein.  
KW Human: endothelial cell proliferation inhibitor; collagen XVII;  
KW angiogenesis inhibitor; anti-tumour; cytostatic; antiproliferative;  
KW vasotropic; dermatological; ophthalmological; vulvarety;  
KW antiarteriosclerotic; antidiabetic; haemostatic; contractile;  
KW ocular angiogenic disease; atherosclerosis; scleroderma;  
KW myocardial angiogenesis; telangiectasia; angiofibroma;  
wound granulation.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200026368-A2.  
XX  
PD 11-MAY-2000.  
XX  
XX  
PF 01-NOV-1999; 99WO-US25605.  
XX  
PR 30-OCT-1998; 98US-0106343.  
XX  
PR 20-MAY-1999; 99US-0315689.  
XX  
XX  
PA (CHIL-) CHILDRENS MEDICAL CENT.  
XX  
XX  
PI O'Reilly MS, Folkman MJ;  
XX  
XX  
DR WPI; 2000-365617/31.

DR N-PSDB: AAA27005.

XX Novel endostatin capable of inhibiting endothelial cell proliferation  
 PT and angiogenesis, useful for treating angiogenesis-dependent cancers  
 PT and as birth control agents

PS Claim 3; Page 39; 68pp; English.

XX The present sequence is an alternate functional endostatin  
 CC protein. When the human endostatin gene sequence AAA27004 is  
 CC recombinantly expressed, an observable doublet of protein results, both  
 CC versions of which are functional endostatin proteins. The present  
 CC endostatin variant is the same as the protein encoded by AAA27004 minus  
 CC the first four amino acids. Recombinant mouse endostatin (20 mg/kg) was  
 CC administered subcutaneously to mice implanted with Lewis lung carcinomas.  
 CC There was tumour mass regression non-detectable levels after 12 days of  
 CC therapy due to the angiogenesis inhibitory activity of endostatin. Thus  
 CC the protein is useful for treatment of angiogenesis-dependent cancers.  
 CC The polynucleotide and polypeptide sequences of this endostatin are  
 CC useful for treating and diagnosis of tumours, ocular angiogenic  
 CC diseases, Osler-Weber syndrome, myocardial angiogenesis, plaque  
 CC neovascularisation, telangiectasia, haemophilic joints, angiolipoma  
 CC and wound granulation, for treatment of diseases related to excessive or  
 CC abnormal stimulation of endothelial cells e.g. intestinal adhesions,  
 CC atherosclerosis, scleroderma. The protein may also be useful as a birth  
 CC control agent by reducing or preventing uterine vascularisation. The  
 CC gene for endostatin may be isolated from cells or tissue that express  
 CC high levels of endostatin, eg. tumour cells, by generating cDNA from  
 CC mRNA using reverse transcriptase and then amplifying the DNA sequence.

XX Sequence 178 AA:

SQ

Query Match 100.0%; Score 346; DB 21; Length 178;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-39;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIYNLDELFPSEALFSGSEGPLKPGARIFSFQDKDYLRRHPTWPKSWHGS 60  
 DB 60 adraavpiynlkdellfpsealfsgsegpikpgarifsfqdkdylrrhptwqkswhgs 119  
 QY 61 DPN 63  
 DB 120 dpn 122

RESULT 2  
 AAB28399  
 ID AAB28399 standard; Protein: 182 AA.  
 XX  
 AC AAB28399;  
 XX  
 DT 19-FEB-2001 (first entry)  
 XX  
 DE Human endostatin.  
 XX  
 KW Human: endostatin; cytostatic; antiproliferative;  
 KW vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;  
 KW cancer; vascularised solid tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200064946-A2.  
 XX  
 PD 02-NOV-2000.  
 XX  
 PF 28-APR-2000; 2000WO-US11367.  
 XX  
 PR 28-APR-1999; 99US-0131432.  
 XX  
 PA (TEXA) UNIV TEXAS SYSTEM.  
 XX  
 PI Thorpe PE, Brekken RA;

XX WPI: 2000-687317/67.

XX Immunogenic composition for the treatment and diagnosis of cancer  
 PT comprises an anti-VEGF (vascular endothelial growth factor) antibody  
 PT binding the same epitope as the monoclonal antibody ATCC PTA 1595 -  
 XX  
 PS Example 10; Page 291-292; 298pp; English.

XX The present invention relates to anti-Vascular Endothelial Growth Factor  
 CC (VEGF) antibodies that bind to the same epitope as the monoclonal  
 CC antibody ATCC PTA 1595 and which significantly inhibit VEGF binding to  
 CC the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF  
 CC receptor VEGFR1. The present sequence is human endostatin. Endostatin  
 CC may be conjugated onto the anti-VEGF antibodies of the present invention.  
 CC The anti-VEGF antibodies of the present invention are useful for the  
 CC treatment and diagnosis of cancer, especially vascularised solid tumours.

XX Sequence 182 AA:

Query Match 100.0%; Score 346; DB 21; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-39;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIYNLDELFPSEALFSGSEGPLKPGARIFSFQDKDYLRRHPTWPKSWHGS 60  
 DB 64 adraavpiynlkdellfpsealfsgsegpikpgarifsfqdkdylrrhptwqkswhgs 123  
 QY 61 DPN 63  
 DB 124 dpn 126

RESULT 3  
 AAY94323  
 ID AAY94323 standard; Protein: 182 AA.

XX  
 AC AAY94323;  
 XX  
 DT 11-AUG-2000 (first entry)  
 XX  
 DE Human endostatin protein.  
 XX  
 KW Human: endothelial cell proliferation inhibitor; collagen XVIII;  
 KW angiogenesis inhibitor; anti-tumour; cytostatic; antiproliferative;  
 KW vasotropic; dermatological; ophthalmological; vulvar;  
 KW antiatherosclerotic; antidiabetic; haemostatic; contraceptive;  
 KW ocular angiogenic disease; atherosclerosis; scleroderma;  
 KW myocardial angiogenesis; telangiectasia; angiolipoma;  
 KW wound granulation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200026368-A2.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 01-NOV-1999; 99WO-US25605.  
 XX  
 PR 30-OCT-1998; 98US-0106343.  
 PR 20-MAY-1999; 99US-0315689.  
 XX  
 PA (CHIL-) CHILDRENS MEDICAL CENT.  
 XX  
 PI O'Reilly MS, Folkman MJ;  
 XX  
 DR WPI: 2000-365617/31.  
 DR N-PSDB: AAA27004.  
 XX  
 PT Novel endostatin capable of inhibiting endothelial cell proliferation  
 PT and angiogenesis, useful for treating angiogenesis-dependent cancers  
 PT and as birth control agents

XX Claim 2: Page 38; 68pp; English.  
 PS  
 CC The present sequence is an endostatin protein which is the carboxy  
 CC terminal protein of human collagen XVIII. Recombinant mouse endostatin  
 CC (20 mg/kg) was administered subcutaneously to mice implanted with Lewis  
 CC lung carcinomas. There was tumour mass regression non-detectable levels  
 CC after 12 days of therapy due to the angiogenesis inhibitory activity of  
 CC endostatin. Thus the protein is useful for treatment of angiogenesis-  
 CC dependent cancers. The polynucleotide and polypeptide sequences of this  
 CC endostatin are useful for treating and diagnosis of tumours, ocular  
 CC angiogenic diseases, Osler-Weber syndrome, myocardial angiogenesis,  
 CC plaque neovascularisation, telangiectasia, haemophilic joints,  
 CC angiofibroma and wound granulation, for treatment of diseases related to  
 CC excessive or abnormal stimulation of endothelial cells e.g. intestinal  
 CC adhesions, atherosclerosis, scleroderma. The protein may also be useful  
 CC as a birth control agent by reducing or preventing uterine  
 CC vascularisation. The gene for endostatin may be isolated from cells or  
 CC tissue that express high levels of endostatin, eg. tumour cells, by  
 CC generating cDNA from mRNA using reverse transcriptase and then amplifying  
 CC the DNA sequence.  
 CC  
 SQ Sequence 182 AA:  
 Query Match 100.0%; Score 346; DB 21; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-39;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ADRAAPVIVNLKDELLFPSEALFSGSEGLPKPGARIFSGDKDVLRHPTWPKSVWHS 60  
 DB 64 adraavpivnlkdeellfpsealifsgseglpkpgarifsfgdkdvlrhptwpskswhs 123  
 QY 61 DPN 63  
 DB 124 dpn 126  
 Db 124 dpn 126  
 RESULT 4  
 AAY59622 standard; protein: 182 AA.  
 ID AAY59622;  
 AC AAY59622;  
 DT 14-MAR-2000 (first entry)  
 DE Human endostatin protein fragment.  
 XX  
 KW Endostatin; scatter factor activity; human; tubulogenesis; psoriasis;  
 KW metastatic cancer; tumorigenesis; ocular angiogenic disease;  
 KW rheumatoid arthritis; Osler-Weber syndrome; telangiectasia;  
 KW haemophilic joint; angiofibroma; wound granulation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9962944-A2.  
 PD 09-DEC-1999.  
 PF 03-JUN-1999; 99WO-US12278.  
 XX  
 PR 03-JUN-1998; 98US-0087890.  
 PR 10-JUL-1998; 98US-0092393.  
 PR 01-SEP-1998; 98US-0098790.  
 XX  
 PA (CHIL-) CHILDRENS MEDICAL CENT.  
 PI Javaherian K, Folkman MJ;  
 DR WPI: 2000-072833/06.  
 PT New endostatin oligomers, used for treating e.g. tumours -  
 XX

PS Disclosure: Page 6; 44pp; English.  
 XX  
 CC This sequence is a fragment of the human endostatin protein. Endostatin  
 CC is an approximately 20kD C-terminal globular domain of the collagen-like  
 CC protein collagen XVIII. Protein oligomers consisting of more than one  
 CC endostatin monomer have anti-tubulogenic effects and induce  
 CC reorganization of the actin cytoskeleton. The oligomer has scatter factor  
 CC activity. The oligomers induce the destruction of tubular lumens and  
 CC elongation of cells, and inhibit tubulogenesis and tumorigenesis. The  
 CC oligomers can also be used to treat metastatic cancers, tumours,  
 CC rheumatoid arthritis, psoriasis, ocular angiogenic disease, Osler-Weber  
 CC syndrome, plaque neovascularisation, telangiectasia, haemophilic  
 CC joints, angiofibroma and wound granulation. The oligomers can also be  
 CC used to treat diseases that have angiogenesis as a pathological  
 CC consequence e.g. ulcers. The endostatin oligomers can also be used to  
 CC develop affinity columns for isolating antibodies or receptors. Passive  
 CC antibody therapy using antibodies that specifically bind endostatin  
 CC oligomers can be used to modulate morphogenic processes such as  
 CC metastatic cancer as well as angiogenesis-dependent processes such as  
 CC reproduction, development, wound healing, tissue repair, and  
 CC angiogenesis-dependent diseases. Also, antisera directed to the Fab  
 CC regions of endostatin oligomer antibodies can be administered to block  
 CC the ability of endogenous endostatin oligomer antisera to bind endostatin  
 CC oligomers.  
 CC  
 SQ Sequence 182 AA:  
 Query Match 100.0%; Score 346; DB 21; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-39;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ADRAAPVIVNLKDELLFPSEALFSGSEGLPKPGARIFSGDKDVLRHPTWPKSVWHS 60  
 DB 64 adraavpivnlkdeellfpsealifsgseglpkpgarifsfgdkdvlrhptwpskswhs 123  
 QY 61 DPN 63  
 DB 124 dpn 126  
 Db 124 dpn 126  
 RESULT 5  
 AAY08693 standard; protein: 183 AA.  
 ID AAY08693;  
 AC AAY08693;  
 DT 10-AUG-1999 (first entry)  
 DE Human endostatin protein fragment.  
 XX  
 KW Plasmidogen; human; angiotensin; endostatin; gene therapy; vector;  
 KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;  
 KW tumour growth; solid tumour; diabetic retinopathy; retina.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9926480-A1.  
 PD 03-JUN-1999.  
 PF 20-NOV-1998; 98WO-US24950.  
 PR 20-NOV-1997; 97US-0975424.  
 XX  
 PA (GENE-) GENETIX PHARM INC.  
 PA (MAST ) MASSACHUSETTS INST TECHNOLOGY.  
 PI Bachelot T, Leboulch P, Pawliuk RJ;  
 DR WPI: 1999-357696/30.  
 DR N-PSDB; AAX77719.  
 XX

PT Anti-angiogenic gene therapy vectors  
XX  
XX Disclosure: Page 74-75; 83pp; English.  
XX  
CC This invention describes a novel viral gene therapy vector comprising a  
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen  
CC from human or murine angiostatin, human or murine endostatin and  
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is  
CC sufficiently attenuated for use in human gene therapy. The products of  
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and  
CC ophthalmological activity. The vector is used in gene therapy for  
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector  
CC expresses an anti-angiogenic polypeptide. An additional use comprises  
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide  
CC inhibits angiogenesis in the vicinity of the retina. The vector is  
CC administered to cells ex vivo and then administered to the patient.  
XX  
SQ Sequence 183 AA:  
  
Query Match 100.0%; Score 346; DB 20; Length 183;  
Best Local Similarity 100.0%; Pred. No. 5.7e-39;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 ADRAAVPIVNLKDELLFPSEWALFSGSEGPLKPGARIFSGDKDYLRIHTPQKSYMHGS 60  
Db 64 adraavpivnlkdellfpsewalfsgseglpkpgarlfsgdkdylrhtpqpksywhgs 123  
QY 61 DPN 63  
Db 124 dpn 126  
  
RESULT 6  
ID AAY02113 standard; Protein: 183 AA.  
XX  
XX AAY02113:  
XX  
XX 16-JUL-1999 (first entry)  
XX  
XX SEQ ID 76 of W09916889.  
XX  
XX Angiostatin; endostatin; Interferon; thrombospondin;  
XX  
XX Interferon-inducible protein; platelet factor 4; anti-angiogenic;  
XX  
XX anti-tumor; multifunctional protein; angiogenesis-mediated disease;  
XX  
XX cancer; diabetic retinopathy; macular degeneration; arthritis;  
XX  
XX tumor cell production.  
XX  
XX  
XX Homo sapiens.  
XX  
XX W09916889-A1.  
XX  
XX 08-APR-1999.  
XX  
XX 30-SEP-1998; 98WO-US20464.  
XX  
XX 01-OCT-1997; 97US-0060609.  
XX  
XX (SEAR ) SEARLE & CO G D.  
XX  
XX Bolanowski MA, Caparon MH, Casperson GF, Gregory SA;  
XX  
XX Klein BK, McKearn JP;  
XX  
XX WPI: 1999-255098/21.  
XX  
XX  
XX New multifunctional proteins useful for treating angiogenic-mediated  
XX  
XX diseases  
XX  
XX Disclosure: Page 106-107; 121pp; English.  
XX  
XX The specification describes multifunctional proteins which comprise  
XX  
XX combinations of angiostatin, endostatin, interferon, thrombospondin,  
XX  
XX

CC Interferon-inducible protein and platelet factor 4, and have  
CC anti-angiogenic and/or anti-tumor activity. The multifunctional protein  
CC may exhibit useful properties such as having similar or greater  
CC biological activity when compared to a single factor or by having  
CC improved half-life or decreased adverse side effects, or a combination  
CC of these properties. The proteins can be used for treating an  
CC angiogenic-mediated disease, e.g. cancer, diabetic retinopathy, macular  
CC degeneration, or arthritis. They can also be used for inhibiting the  
CC production of tumor cells (characteristic of lung, breast, ovarian,  
CC prostate, pancreatic, gastric, colon, renal, bladder cancers; melanoma,  
CC hepatoma, sarcoma and lymphoma) in a patient and for inhibiting tumor  
XX  
XX growth. The present sequence is used in the course of the invention.  
XX  
SQ Sequence 183 AA:  
  
Query Match 100.0%; Score 346; DB 20; Length 183;  
Best Local Similarity 100.0%; Pred. No. 5.7e-39;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 ADRAAVPIVNLKDELLFPSEWALFSGSEGPLKPGARIFSGDKDYLRIHTPQKSYMHGS 60  
Db 64 adraavpivnlkdellfpsewalfsgseglpkpgarlfsgdkdylrhtpqpksywhgs 123  
QY 61 DPN 63  
Db 124 dpn 126  
  
RESULT 7  
ID AAB30493 standard; Protein: 183 AA.  
XX  
XX AAB30493:  
XX  
XX 06-MAR-2001 (first entry)  
XX  
XX  
XX Amino acid sequence of human endostatin encoded by plasmid pMALCH#15.  
XX  
XX Streptomyces sp. strain C5; SnPA; S. venezuelae; alpha-amylase;  
XX  
XX endostatin; cancer; tumour growth; angiogenesis.  
XX  
XX  
XX Homo sapiens.  
XX  
XX W0200060945-A1.  
XX  
XX 19-OCT-2000.  
XX  
XX 12-APR-2000; 2000WO-US09747.  
XX  
XX 13-APR-1999; 99US-0129084.  
XX  
XX (MERI ) MERCK & CO INC.  
XX  
XX Desanti CL, Strohl WR;  
XX  
XX WPI: 2000-686970/67.  
XX  
XX N-PSDB; AAC62023.  
XX  
XX Preparation of soluble recombinant endostatin involves transforming  
XX  
XX Streptomyces host with expression vector comprising nucleotide  
XX  
XX sequence encoding endostatin operably linked to linker and leader  
XX  
XX peptide  
XX  
XX  
XX Example 1: Fig 6; 57pp; English.  
XX  
XX The present sequence represents human endostatin. The protein is  
XX  
XX expressed in Streptomyces. Leader sequences of Streptomyces sp. strain  
XX  
XX C5 SnPA and S. venezuelae alpha-amylase proteins are linked to the  
XX  
XX N-terminal of endostatin. This ensures that endostatin protein is  
XX  
XX produced as a secreted, soluble protein which needs no refolding, is  
XX  
XX stable in the fermentation broth and is produced in large quantities.  
XX  
XX The method is used for preparing soluble recombinant human, murine or

CC primate endostatin, which is useful in the treatment of cancer,  
CC inhibition of tumour growth, inhibition of angiogenesis, isolation of  
CC receptors for endostatin and for identification of anti-angiogenic  
CC compounds in assays. The endostatin protein is produced as a secreted,  
CC soluble protein which needs no refolding, is stable in the fermentation  
CC broth and is produced in large quantities. Streptomycetes are amenable  
CC for cultivation in large fermentations allowing for large quantities of  
CC soluble endostatin to be produced.  
XX  
SQ Sequence 183 AA:

Query Match 100.0%; Score 346; DB 21; Length 183;  
Best Local Similarity 100.0%; Pred. No. 5,7e-39;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELFPSEALFSGSEGPLKPGARIFSFDDKDVLRHPTWQKSYMHGS 60  
DB 64 adraavpivnlkdellfpsealifsgseglpkparifsfddgkdvlrhptwqkswmhs 123  
QY 61 DPN 63  
DB 124 dpn 126

## RESULT 8

AAB16451  
ID AAB16451 standard; Protein; 183 AA.

AC AAB16451;

DT 27-OCT-2000 (first entry)

DE Human endostatin protein sequence.

KM Angiogenesis-inhibiting protein receptor; angiogenesis; angiotatin;  
KM endostatin; plasminogen; laminin; treatment; wound healing; solid tumour;  
KM psoriasis; scleroderma; myocardial angiogenesis; Crohn's disease;  
KM cerebral collateral; arteriovenous malformation; rheobosis; cancer;  
KM diabetic retinopathy; arthritis; wound healing; peptic ulcer;  
KM Helicobacter related disease; fracture; cat scratch fever.

XX Homo sapiens.

OS WO200032631-A2.

PN 08-JUN-2000.

PD 06-DEC-1999; 99WO-US28897.

PF 04-DEC-1998; 98US-0206059.

PR (ENTR-) ENTREMED INC.

PA Macdonald NJ, Sim KL;

PI Macdonald NJ, Sim KL;

PT WPI; 2000-412290/35.

PT New angiogenesis-inhibiting protein receptors, useful in methods for  
PT treating diseases and processes that are mediated by angiogenesis -  
PT as solid tumours, psoriasis, scleroderma and myocardial angiogenesis -

PS Disclosure; Figure 3; 100pp; English.

XX This invention relates to angiogenesis-inhibiting protein receptors, and  
XX the DNA sequences encoding them. Angiogenesis is the generation of new  
XX blood vessels into a tissue, and normally occurs in wound healing,  
XX foetal and embryonal development and the formation of the corpus luteum,  
XX endometrium and placenta. Angiostatin is a protein (see AAB16450 and  
XX AAA68202) involved in angiogenesis, and has an amino acid sequence  
XX similar to that of a plasminogen fragment (see murine plasminogen  
XX AAB16490). Angiostatin has the ability to inhibit angiogenesis.  
CC Endostatin is also an angiogenesis inhibiting protein (see AAB16451 and

CC AAA68203). Sequences AAA68242 and AAB16522 represent coding and protein  
CC sequences of human laminin. Laminin is an angiotatin binding protein,  
CC and some of the peptides of the invention share homology with regions of  
CC laminin. Peptides AAB16452-B16521 (excluding AAB16490) are the  
CC angiogenesis-inhibiting protein receptor fragments of the invention. The  
CC peptides bind either angiotatin or endostatin and can be used in methods  
CC for treating diseases and processes that are mediated by angiogenesis,  
CC such as solid tumours, psoriasis, scleroderma, myocardial angiogenesis,  
CC Crohn's disease, cerebral collaterals, arteriovenous malformations,  
CC rheobosis, diabetic retinopathy, arthritis, wound healing, peptic ulcers,  
CC Helicobacter related diseases, fractures, placentaion and cat scratch  
CC fever. They are useful for the detection and prognosis of cancer. DNA  
CC sequences A628204-A628241 encode the peptides of the invention.  
XX  
SQ Sequence 183 AA:

Query Match 100.0%; Score 346; DB 21; Length 183;  
Best Local Similarity 100.0%; Pred. No. 5,7e-39;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELFPSEALFSGSEGPLKPGARIFSFDDKDVLRHPTWQKSYMHGS 60  
DB 64 adraavpivnlkdellfpsealifsgseglpkparifsfddgkdvlrhptwqkswmhs 123

QY 61 DPN 63  
DB 124 dpn 126

## RESULT 9

AA90771  
ID AA90771 standard; Protein; 183 AA.

AC AA90771;

DT 22-AUG-2000 (first entry)

DE Human angiogenesis inhibiting factor 1 protein.

KM Human; angiogenesis inhibiting factor 1; IAF-1; tumour; antibody;  
KM abnormal vessel disease.

XX Homo sapiens.

OS CN1244536-A.

PN 16-FEB-2000.

PD 10-AUG-1998; 98CN-0117150.

PF 10-AUG-1998; 98CN-0117150.

PR (ONCO-) INST ONCOLOGY UNDER TUMOR HOSPITAL CHINE.

PA Yang Z, Guo W;

PI Yang Z, Guo W;

PT WPI; 2000-388168/34.

PT N-PSDB; AAA29884.

PT Angiogenesis inhibiting factor 1 and its derivative useful for treating  
PT tumors -

PS Claim 1; Fig 5; 41pp; Chinese.

XX The present sequence represents an angiogenesis inhibiting factor (1),  
XX designated IAF-1. The present invention also describes: (1) preparation  
XX of (1) and its derivative; (2) an IAF binding acceptor and its  
XX preparation; and (3) an IAF antibody. (1) is useful for preparing new  
XX biological preparations for effectively treating various tumours and  
XX abnormal vessel diseases. The IAF antibody is preferably a polyclonal  
XX antibody, mosaic antibody, single stranded antibody and human originated  
XX antibody.

XX Sequence 183 AA;  
 SQ  
 Query Match 100.0%; Score 346; DB 21; Length 183;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-39;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELFPSEALFSGSEGPLKPGARIFSGDKDVLRRHPTWPKSWHGS 60  
 |||  
 DB 64 adraavpivnlkdellfpsealfsgseglkpgarifsfqdkdvlrrhptwqkswwhgs 123  
 QY 61 DPN 63  
 |||  
 DB 124 dpn 126

RESULT 10  
 AAY70252  
 ID AAY70252 standard; Protein; 183 AA.  
 \*XX AAY70252;  
 AC  
 XX  
 XX  
 DT 06-JUN-2000 (first entry)  
 XX  
 DE Human angiogenesis inhibitor, endostatin.  
 XX  
 XX Human; immunoglobulin gamma Fc fragment; endostatin; immunofusin;  
 KW angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritis;  
 KW antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;  
 KW vasoregic; vulnerary; treatment; arteriosclerosis; tumor;  
 KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;  
 KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;  
 KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;  
 KW wound granulation; keloid scar; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200011033-A2.  
 XX  
 PD 02-MAR-2000.  
 XX  
 PE 25-AUG-1999; 99WO-US19329.  
 XX  
 PF 25-AUG-1998; 98US-0097883.  
 \*PR 25-AUG-1998; 98US-0097883.  
 XX  
 PA (LEXI-) LEXINGEN PHARM CORP.  
 XX  
 PI Lo K, Li Y, Gillies SD;  
 XX  
 DR WPI: 2000-237616/20.  
 DR N-PSDB: AA251291.  
 XX  
 PT Novel fusion protein of angiotensin or endostatin and an immunoglobulin  
 PT FC region, useful for treating conditions mediated by angiogenesis,  
 PT such as rheumatoid arthritis, tumors and macular degeneration -  
 XX  
 XX Example 1; Pages 41-42; 68pp; English.

The patent discloses a DNA molecule encoding a fusion protein comprising a signal sequence, an immunoglobulin Fc region, and an angiogenesis inhibitor selected from angiotensin, endostatin, a plasminogen fragment having angiotensin activity, a collagen XVIII fragment having endostatin activity, or combinations of them. The fusion protein (immunofusin) is used to inhibit angiogenesis and to treat diseases or conditions mediated by angiogenesis. Conditions that may be treated include solid tumors, blood born tumors, tumor metastasis, benign tumors including hemangiomas, acoustic neuromas, neurofibromas, trachomas and pyogenic granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases e.g. diabetic retinopathy, retinopathy of prematurity, macular degeneration, corneal graft rejection, neovascular glaucoma, retrolental fibroplasia, rubecis and Osler-Weber syndrome; myocardial angiogenesis, plaque neovascularisation, telangiectasia, haemophilic joints'

CC angiofibroma, wound granulation, and excessive or abnormal stimulation of  
 CC endothelial cells, intestinal cells, atherosclerosis, sclerodermal and  
 CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used  
 CC in gene therapy. The present sequence is a human endostatin used in the  
 CC construction of immunofusin containing human immunoglobulin gamma  
 CC (IgG) Fc fragment.  
 XX  
 SQ Sequence 183 AA;  
 Query Match 100.0%; Score 346; DB 21; Length 183;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-39;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELFPSEALFSGSEGPLKPGARIFSGDKDVLRRHPTWPKSWHGS 60  
 |||  
 DB 64 adraavpivnlkdellfpsealfsgseglkpgarifsfqdkdvlrrhptwqkswwhgs 123  
 QY 61 DPN 63  
 |||  
 DB 124 dpn 126

RESULT 11  
 AAB49379  
 ID AAB49379 standard; Protein; 183 AA.  
 XX  
 AC AAB49379;  
 XX  
 DT 02-MAR-2001 (first entry)  
 XX  
 DE Human endostatin SEQ ID NO: 2.  
 XX  
 KW Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;  
 KW cancer; inflammation; angiogenesis-dependent disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200067771-A1.  
 XX  
 PD 16-NOV-2000.  
 XX  
 PE 02-MAY-2000; 2000WO-US12063.  
 XX  
 PF 06-MAY-1999; 99US-0132907.  
 PR 14-JUL-1999; 99US-0353333.  
 XX  
 PA (BURN-) BURNHAM INST.  
 XX  
 PI Vuori K;  
 XX  
 DR WPI: 2001-040937/05.  
 DR N-PSDB: AAC88285.  
 XX  
 PT Endostatin peptide comprising at least four endostatin amino acid  
 PT residues are e.g. angiogenesis inhibitors for treating cancer and  
 PT diabetic retinopathy -  
 XX  
 XX Disclosure; Fig 1; 146pp; English.

The present invention provides endostatin peptides which can be used in the modulation of angiogenesis. This is useful in the treatment of cancers, inflammation, rheumatoid arthritis, chronic articular rheumatism, psoriasis, disorders associated with inopportune invasion of vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy of prematurity, macular degeneration, corneal graft rejection, retrolental fibroplasia, rubecis, capillary proliferation in atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent diseases include Osler-Weber syndrome, myocardial angiogenesis, plaque neovascularisation, telangiectasia, haemophilic joints and wound granulation. In addition, the peptides can be used as birth control agents.

SQ Sequence 183 AA;

Query Match 100.0%; Score 346; DB 22; Length 183;  
Best Local Similarity 100.0%; Pred. No. 5.7e-39;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFPSEALFSGSGEPLKPGARIFSPDGKDVLRHPTWPKSVWHS 60  
Db adraavpivnlkdelldfswaalfsgseplkpgarilfsfdgkdvlrhptwpskswvhs 123  
|||

QY 61 DPN 63  
Db 124 dpn 126

RESULT 12  
AAW90877  
ID AAW90877 standard; protein: 193 AA.  
.XX  
AC AAW90877;  
DT 07-JUL-2000 (first entry)  
XX  
\*DE Human HMW endostatin (4) protein.  
XX  
KM Endostatin; human; renal insufficiency; antitumor; antiproliferative;  
KM treatment; angiogenesis; tumor; vascular disease.  
XX  
OS Homo sapiens.  
XX  
PN MO200017240-A1.  
PD 30-MAR-2000.  
XX  
PF 21-SEP-1999; 99WO-EP06963.  
XX  
PR 21-SEP-1998; 98DE-1042992.  
PR 03-APR-1999; 99DE-1015267.  
PR 08-JUN-1999; 99DE-1026040.  
XX  
PA (HAEM-) HAEMOPEP PHARMA GMBH.  
XX  
PI Staendker L, Forssmann W;  
XX  
DR WPI: 2000-292826/25.  
XX  
PT New high molecular weight form of endostatin, useful e.g. as  
PT antiangiogenic agent for treating cancer, isolated from hemofiltrate of  
PT patients with kidney failure -  
XX  
PS Claim 2; Page 21-22; 32pp; German.  
XX  
CC This invention describes novel human high molecular weight (HMW)  
CC endostatin (he) protein fragments isolated from the hemofiltrate of  
CC patients with renal insufficiency. The products of the invention have  
CC antitumor and antiproliferative activity. he is used to treat; (i)  
CC diseases that involve uncontrolled angiogenesis, particularly tumors;  
CC and (ii) vascular diseases of supporting or connective tissue;  
CC respiratory tract, cardiovascular system, urogenital tract and nervous  
CC system, or sensory organs (particularly the eye). he is also used to  
CC raise specific antibodies which are used for diagnosis and treatment of  
CC conditions that involve overexpression of he. he has a very long plasma  
CC half-life and can be administered repeatedly without inducing an immune  
CC response. AAW90874-W90877 represent the endostatin proteins described in  
CC the method of the invention.  
XX  
SQ Sequence 193 AA;

Query Match 100.0%; Score 346; DB 21; Length 193;  
Best Local Similarity 100.0%; Pred. No. 6.1e-39;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFPSEALFSGSGEPLKPGARIFSPDGKDVLRHPTWPKSVWHS 60  
Db 77 adraavpivnlkdelldfswaalfsgseplkpgarilfsfdgkdvlrhptwpskswvhs 136  
|||

QY 61 DPN 63  
Db 137 dpn 139

RESULT 13  
AAW90874  
ID AAW90874 standard; protein: 195 AA.  
XX  
AC AAW90874;  
DT 07-JUL-2000 (first entry)  
XX  
DE Human HMW endostatin (1) protein.  
XX  
KM Endostatin; human; renal insufficiency; antitumor; antiproliferative;  
KM treatment; angiogenesis; tumor; vascular disease.  
XX  
OS Homo sapiens.  
XX  
PN MO200017240-A1.  
PD 30-MAR-2000.  
XX  
PF 21-SEP-1999; 99WO-EP06963.  
XX  
PR 21-SEP-1998; 98DE-1042992.  
PR 03-APR-1999; 99DE-1015267.  
PR 08-JUN-1999; 99DE-1026040.  
XX  
PA (HAEM-) HAEMOPEP PHARMA GMBH.  
XX  
PI Staendker L, Forssmann W;  
XX  
DR WPI: 2000-292826/25.  
XX  
PT New high molecular weight form of endostatin, useful e.g. as  
PT antiangiogenic agent for treating cancer, isolated from hemofiltrate of  
PT patients with kidney failure -  
XX  
PS Claim 2; Page 18-19; 32pp; German.  
XX  
CC This invention describes novel human high molecular weight (HMW)  
CC endostatin (he) protein fragments isolated from the hemofiltrate of  
CC patients with renal insufficiency. The products of the invention have  
CC antitumor and antiproliferative activity. he is used to treat; (i)  
CC diseases that involve uncontrolled angiogenesis, particularly tumors;  
CC and (ii) vascular diseases of supporting or connective tissue,  
CC respiratory tract, cardiovascular system, urogenital tract and nervous  
CC system, or sensory organs (particularly the eye). he is also used to  
CC raise specific antibodies which are used for diagnosis and treatment of  
CC conditions that involve overexpression of he. he has a very long plasma  
CC half-life and can be administered repeatedly without inducing an immune  
CC response. AAW90874-W90877 represent the endostatin proteins described in  
CC the method of the invention.  
XX  
SQ Sequence 195 AA;

Query Match 100.0%; Score 346; DB 21; Length 195;  
Best Local Similarity 100.0%; Pred. No. 6.2e-39;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFPSEALFSGSGEPLKPGARIFSPDGKDVLRHPTWPKSVWHS 60  
Db 77 adraavpivnlkdelldfswaalfsgseplkpgarilfsfdgkdvlrhptwpskswvhs 136  
|||

QY 61 DPN 63

Db 137 dpn 139

## RESULT 14

AAB30495  
ID AAB30495 standard; protein: 216 AA.

AC AAB30495;

DT 06-MAR-2001 (first entry)

DE Amino acid sequence of vaa-endostatin fusion protein in PANT3052.

XX Streptomyces sp. strain C5; SnpA; S. venezuelae; alpha-amylase;

KW endostatin; cancer; tumour growth; angiogenesis.

XX Synthetic.

OS Streptomyces sp.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..28

FT /note= "vaa signal sequence"

FT Protein 29..216

WT /note= "endostatin"

XX WO200060945-A1.

PN 19-OCT-2000.

PD 12-APR-2000; 2000WO-US09747.

XX 13-APR-1999; 99US-0129084.

XX (MERI ) MERCK & CO INC.

XX Desantl CL, Strohl WR;

XX WPI: 2000-686970/67.

DR N-PSDB: AAC62025.

XX Preparation of soluble recombinant endostatin involves transforming

PT Streptomyces host with expression vector comprising nucleotide

PT sequence encoding endostatin operably linked to linker and leader

PT peptide

XX Example 1; Fig 10A-B; 57pp; English.

PS The present sequence represents a fusion protein of vaa and endostatin.

XX The specification describes a method for the production of soluble,

CC recombinant human endostatin in streptomyces. leader sequences of

CC Streptomyces sp. strain C5 SnpA and S. venezuelae alpha-amylase proteins

CC are linked to the N-terminal of endostatin. This ensures that endostatin

CC protein is produced as a secreted, soluble protein which needs no

CC refolding, is stable in the fermentation broth and is produced in large

CC quantities. The method is used for preparing soluble recombinant human,

CC murine or primate endostatin, which is useful in the treatment of cancer,

CC inhibition of tumour growth, inhibition of angiogenesis, isolation of

CC receptors for endostatin and for identification of anti-angiogenic

CC compounds in assays. The endostatin protein is produced as a secreted,

CC soluble protein which needs no refolding, is stable in the fermentation

CC broth and is produced in large quantities. Streptomyces are amenable

CC for cultivation in large fermentations allowing for large quantities of

CC soluble endostatin to be produced.

XX Sequence 216 AA;

SQ

Query Match 100.0%; Score 346; DB 21; Length 216;

Best Local Similarity 100.0%; Pred. No. 7, 1e-39;

Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELPSPWEALFSGSEGPLKPGCARIFSPDGKDYLRHPMPQKSVNHGS 60  
|||||  
Db 97 adraavpivnlkdeallfswaellfsgsegrlpkparlfsfsgkvdlrhpmpqkswvhgs 156

QY 61 DPN 63  
|||  
Db 157 dpn 159

## RESULT 15

AAW26327  
ID AAW26327 standard; protein: 684 AA.

AC AAW26327;

DT 19-NOV-1997 (first entry)

DE Human alpha-1 collagen (XVIII).

XX Alpha-1 collagen; type XVIII collagen; cartilage degeneration.

XX Homo sapiens.

XX Key

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

FT Peptide

Location/Qualifiers  
1..6  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
7..12  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
13..18  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
19..24  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
25..30  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
31..36  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
37..42  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
43..48  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
49..54  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
55..59  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
60..65  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
66..71  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
72..77  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
78..83  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
84..89  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
90..95  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
96..101  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
102..107  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
108..113  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
114..119  
/label= GXYGX'Y' motif  
/note= "Claim 1"  
120..126



```
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 129..134
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 135..140
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 141..146
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 147..152
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 153..158
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 159..164
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 165..170
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 171..176
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 181..186
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 187..192
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 193..198
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 215..220
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 221..226
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 227..232
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 233..238
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 239..244
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 257..262
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 263..268
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 269..274
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 275..280
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 286..291
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 292..297
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 298..303
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 309..314
FT /label= GYXGX'Y' motif
```

```
FT /note= "Claim 1"
FT Peptide
FT 315..320
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 322..328
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 329..334
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 335..340
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 354..359
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 360..365
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 366..372
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 523..528
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 542..547
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"
FT Peptide
FT 590..595
FT /label= GYXGX'Y' motif
FT /note= "Claim 1"

US5643783-A.
PN 01-JUL-1997.
PD 01-DEC-1993; 930S-0159784.
PR 01-DEC-1993; 930S-0159784.
XX (HARD ) HARVARD COLLEGE.
XX Oh SP, Olsen BR;
XX WIPI: 1997-350247/32.
XX N-PSDB: AAT84484.
XX Nucleic acid encoding human alpha-1 collagen - for production of
XX recombinant alpha-1 collagen, for use in the treatment of cartilage
XX degeneration
XX Claim 1; Column 23-30; 35pp; English.
XX
XX Novel human type alpha-1 (XVIIII) collagen is characterised by
XX 10 triple helical domains containing the GYXGX'Y' motif (where X,
XX Y, X' and Y' represent any amino acid), the helical domains being
XX separated and flanked by non-triple helical regions which may
XX provide flexibility. Alpha-1 collagen is expressed in multiple
XX tissues, especially liver, lung and kidney. A claimed plasmid
XX comprising alpha-1 collagen nucleic acid (see AAT84484) and an
XX expression control sequence can be used to express recombinant
XX collagen in prokaryotic or eukaryotic (especially mammalian) host
XX cells. The alpha-1 collagen may be used to treat a patient
XX suffering from a disease associated with cartilage degradation, and
XX for supplementing collagen. It can also be used as a connective
XX
Query Match 100.0%; Score 346; DB 18; Length 684;
Best Local Similarity 100.0%; Pred. No. 3e-38;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ADRAAPVIYNLKDELFPSEALFSGSEGPLKPGARIFSPDKDVLRHPTWPKSVWHS 60
DB 565 adraapviynlkdelfpsealifsgseplkpgarilfsidgkxvllrhpwqkswngs 624
```

QY 61 DPN 63  
|||  
Db 625 dpn 627

Search completed: September 10, 2001, 13:49:11  
Job time: 106 sec